TWO TOOLS FOR THREE CHARACTERIZATIONS OF CELLS: MACHINE LEARNING FOR AUTOMATED CANCER CELL IDENTIFICATION AND ELECTROROTATION FOR CANCER CELL AND ALGAE CELL ELECTRICAL PROPERTY MEASUREMENT

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Timothy Burke Lannin

January 2017

TWO TOOLS FOR THREE CHARACTERIZATIONS OF CELLS: MACHINE LEARNING FOR AUTOMATED CANCER CELL IDENTIFICATION AND ELECTROROTATION FOR CANCER CELL AND ALGAE CELL ELECTRICAL PROPERTY MEASUREMENT

Timothy Burke Lannin, Ph. D.

Cornell University 2016

Advances in rare cell capture technology have made possible the interrogation of circulating tumor cells (CTCs) captured from patient blood. However, manually locating captured cells in the device bottlenecks data processing by being tedious (hours per sample) and compromises the results by being inconsistent and prone to user bias. In the first aim of my thesis, I employed machine learning algorithms to locate and classify thousands of possible cells in a few minutes rather than a few hours, representing an order of magnitude increase in processing speed performance. Optimal algorithm selection depends on the peculiarities of the individual dataset, indicating the need of a careful comparison and optimization of algorithms for individual image classification tasks.

The capture of circulating tumor cells via immuno-affinity may be compromised by reduced antigen expression associated with acquired resistance to chemotherapy, deprivation of growth factors when entering circulation, and the epithelial-to-mesenchymal transition. Dielectrophoresis (DEP), however, could enhance the capture of these rare cells by attracting cells to the antibody-coated surface. In order to reliably

use dielectrophoresis, cancer cell crossover frequencies must remain lower than those of white blood cells. For my second aim, I used automated electrorotation to measure the cytoplasmic permittivity, cytoplasmic conductivity, and specific membrane capacitance of pancreatic cancer cells under three treatments: 1) acquired resistance to gemcitabine, 2) serum starvation, and 3) induced epithelial-to-mesenchymal transition. I found that the median computed crossover frequency for cancer cells under all treatments remains significantly below that of blood cells, indicating that DEP is a promising technique for enhancing capture.

Algae are a promising feedstock for biofuels, and there is a critical need for a rapid, inexpensive, and label-free measurement of lipid accumulation in algae cells. Measuring the electrical properties of algae has shown promise in the literature for monitoring lipid accumulation because it correlates with a decrease in effective cytoplasmic conductivity. Previous models, however, have assumed a constant cytoplasmic permittivity through the lipid accumulation process, and that assumption must be validated. For my third aim, I used automated electrorotation to measure properties of *Chlamydomonas reinhardtii* cells undergoing lipid accumulation.

BIOGRAPHICAL SKETCH

Tim grew up in Milford, New Hampshire with his older brother, Ben, his mother, Anne, and his father, Danny. He attended the Milford public schools, at which he developed a love for learning under many caring teachers and an appreciation for community with many loyal friends.

In 2007, Tim began his B.S. in Mechanical Engineering at Tufts University. While at Tufts, Tim worked with Dr. Chris Rogers at the Center for Engineering Education and Outreach to develop and teach a course, "Engineering for Non-Engineers," through the Experimental College. He also had the pleasure of conducting undergraduate research on the mechanics of orthopedic sawing with his research adviser, Dr. Tom James.

In 2011, Tim began his doctoral studies at Cornell's Sibley School of Mechanical and Aerospace Engineering under the advisement of Dr. Brian Kirby. He had the pleasure of publishing his work in peer-reviewed journals and presenting at external conferences. His passion for teaching continued at Cornell, where he served as a trainer of teaching assistants and he received an outstanding TA award. The experience of teaching extended beyond the classroom with mentorship of several undergraduate students.

Following his work at Cornell, Tim will join the Mechanical Engineering Department at Lafayette College in Easton, PA, beginning his career in undergraduate engineering education.

Dedicated to everyone whose dedication to me I aspire to match.

ACKNOWLEDGMENTS

First, I would like to thank my advisor, Dr. Brian J. Kirby, for his guidance. I appreciate his support of my goals, both academic and personal. Our success is his success. I would also like to thank my other committee members, Dr. Ankur Singh and Dr. Jerry Feigenson, for their thoughtful comments on my work.

I would also like to thank many research colleagues. My Kirbylab mates have been an unmatched source of knowledge and inspiration. In particular, I would like to thank Fredrik Thege for his friendship and consults in the realms of biomedical engineering and laboratory techniques; Jack Hartman for day-to-day discussions; Mike Bono for his thoughts on electrokinetics and statistical inference and his detailed protocols; Charlie Huang for his help in the CNF; Jim Smith, Steven Santana, Erica Pratt, Alex Barbati, Conor Gruber, Ian Cardle, Nitya Arasanapali, Shalu Suri, and Marie Godla for serving as excellent Kirbylab role models; and Kate Polhemus, Kyle Wellmerling, and Prateek Sehgal for their fresh perspectives and excitement about the future direction of Kirbylab. I owe special thanks to Wey-Wey Su, an immensely talented undergraduate, without whom the last few years of my work would have been far less deep, expedient, and enjoyable. I couldn't have asked for more. I would also like to thank Matt Brendel, Linda Zheng, and Bryan Aristega for the opportunity to mentor them, as well.

Marcia Sawyer deserves countless thanks for dedication to the wellbeing of the students and helping me navigate the administrative side of Cornell. I would also like to acknowledge the efforts of collaborators, faculty, and facilities, including Dr. Andy Rhim; Matt Sung and Dr. Evi Giannakakou at Weill Cornell; Dr. Beth Ahner and Dr. Lubna Richter for their help with algae culture; Dr. Itai Cohen, Brian Leahy, and Lena Bartell for confocal microscopy; Dr. Becky Williams and Carol Bayles in the BRC

Imaging Facility; Dr. Thorsten Joachims and Dr. Kilian Weinberger; Dr. Andy Runa; Dr. Charles Williamson; and the staff at the Cornell Nanoscale Facility.

Thanks to Linda Tompkins and my cohort of TA Trainers, (John Foo, Justin Rosch, Nicole Ricapito, Nicole Hill, Aravind Natarajan, Joseph Carlin Jr, Akanksha Agrawal), for being so formative in my teaching identity.

For laying a foundation prior to Cornell, I would also like to thank my two advisors at Tufts, Dr. Chris Rogers, for his mentorship in undergraduate peer teaching, and Dr. Tom James, for his mentorship in undergraduate research. Additionally, I would like to thank Dr. Vin Manno and Dr. Rob White for their mentorship.

Perhaps most importantly, making it through any part of life, much less my doctoral degree, would have been impossible without the support of my loving friends and family. My mom, Anne Burke, my dad, Danny Lannin, and my older brother, Ben, fostered a love of learning and value of creativity from day one and continued to support me selflessly throughout my education. Thanks also to my dearest Katie; there is nobody else with whom I would rather share a burden or a laugh. I owe my housemates and dear friends Matt Kelly and David Peck a debt for many fond memories. Thanks also to my many other family and friends (Shannon Draucker, Claire Draucker, Matt Gott, Jon Gott, Ty Webster, Jake Potelle, Josh Blase, Nick Kliatsko, Patrick Stanton, Nick Albina, Kendra Potelle, Julia Wagner, Michael Motola-Barnes, Eric Semple, Lorraine Weis, JJ Melfi, Lauren Lazarus, Amy Engelbrecht-Wiggans, Nate Rogers, Nozomi Hitomi, Eddie Bennevie, Ben Reinhardt, Sam Carouso, and Sarah Wengert) who have kept me happy and sane from Milford, Tufts, Cornell, and beyond.

Funding was provided by Cornell University and the National Science Foundation Graduate Research Fellowship Program under Grant No. (DGE-1144153).

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 2

CHAPTER 3

CHAPTER 4

LIST OF TABLES

CHAPTER 1

Introduction

1.1 Introduction

One putative mode of metastasis is by way of circulating tumor cells (CTCs), in which cells detach from a tumor, disseminate through the bloodstream, and form new tumors at distant sites. Kirbylab has developed microfluidic tools to capture and purify CTCs from patient blood. These CTCs are exceedingly rare, of the order of 1-10's of CTCs per mL versus 10^6-10^7 leukocytes per mL, making the capture of pure samples a technical challenge (1). Captured white blood cells can often outnumber CTCs, which necessitates a post-capture analysis to omit blood cells from later analyses. Manual enumeration has been used in the literature (2–9) require expert users to review images of candidate cells; however, manual classification is limited by user fatigue and inter-user variability. Automating the cell identification and classification can make classification repeatable and fast.

In the literature, only minimal exploration has been done in the selection of machine learning method as an element of the design space (10–12). As will be shown in Chapter 2, exploration and thorough optimization of different machine learning algorithms becomes increasingly important when the data is noisy. First, we compared four machine learning methods to more thoroughly explore the design space. Second, we determined whether extracting a set of image features (as done by Scholtens et al.) is superior to putting pixel intensities straight into the algorithms (as done by Svensson

et al. 2015). Third, we explored the effect of training the algorithm on one dataset but testing it on another.

CTCs have been successfully captured from blood samples, but challenges exist capturing subsets of CTCs that have reduced antigen expression. Reduced expression is associated with the epithelial to mesenchymal transition (EMT)(13) or acquired resistance to chemotherapy(14). An improvement to capture could be achieved through dielectrophoresis (DEP). The use of nonuniform alternating current electric fields can be used to attract and repel objects with different complex permittivity values depending on the field frequency. A proof-of-concept hybrid DEP-immunoaffinity devices has shown promise of further improving capture(15–17) and simulations have predicted capture efficiency increases of up to 400% (18). Capture enhancement can be done by operating DEP at an optimal frequency between the crossover frequency of circulating cancer cells and potentially contaminating white blood cells. By measuring an electrorotation spectrum, predict DEP behavior (19–21).

Using cell lines exclusively cannot be a direct substitute for capture models for CTCs in the bloodstream because of possible changes in phenotype in these cells. Changes in CTC size and surface markers can occur due to vesicle shedding processes (22), the epithelial-to-mesenchymal transition (EMT) (4,23,24), or acquired drug resistance (25). Such acquired differences could also result in different DEP spectra, changing crossover frequency. In Chapter 3, we measured electrorotation spectra of gemcitabine-resistant, serum starved, and EMT-induced pancreatic cancer cell lines and healthy donor white blood cells in order to fit cell electrical properties and compute the full effects on DEP spectra and draw conclusions about whether a

crossover frequency separation window exists after the above perturbations are applied.

Finally, algal lipids have potential as a sustainable fuel source (26,27). It has been shown that environmental stresses can induce algal lipid accumulation (26,28), and that in particular, nitrogen starvation of several algae species has been immensely successful(26,29). This process can vary with environmental conditions, so there is therefore a need for real-time lipid measurement to indicate maximum lipid content in order to inform optimal harvest times.

Traditional methods of measuring lipid content have drawbacks, but electrokinetic techniques such as dielectrophoresis (DEP) or impedance cytometry are suitable because these methods can quickly return information about dielectric properties (related to lipid content) of cells nondestrictuviely and sort cells of different compositions. Previous research has shown that cytoplasmic permittivity and/or conductivity is different for high- and low-lipid algae cells(30–33).

There are drawbacks to using these DEP for studying cells. As shown by (34) and discussed in the context of algae by (30,31), DEP upper crossover frequency, depends on both cytoplasmic conductivity and permittivity, so an assumption of constant cytoplasmic permittivity was necessary to estimate conductivity. However, changes in lipid content in algae cells most likely changes both conductivity and permittivity because bulk lipids have both a lower conductivity and permittivity than water. We seek to understand know how both conductivity and permittivity change on a single-cell basis, which necessitates acquiring single-cell, instead of population-averaged, dielectric spectra.

Electrorotation has been used to characterize snow algae (35) as well as for *Chlorella protothecoides* in different nutrient conditions(36). Electrorotation spectra measurements afford a more complete characterization of the electrical property changes with increased lipid content, and allow for better informed design of effective DEP cell sorters and impedance cytometry measurement tools. In Chapter 4, we look to measure electrorotation spectra of both nitrogen replete and nitrogen starved *Chlamydomonas reinhardtii*, and extract single-cell electrical properties, including both conductivity and permittivity, that will allow us to correlate these changes to amounts of lipid accumulation (as measured by BODIPY staining and flow cytometry).

1.2 References

1. Den Toonder J. Circulating tumor cells: the Grand Challenge. Lab Chip 2011;11:375–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21206959. Accessed March 26, 2013.

2. De Bono JS, Scher HI, Montgomery RB, Parker C, Miller MC, Tissing H, Doyle G V, Terstappen LWWM, Pienta KJ, Raghavan D. Circulating tumor cells predict survival benefit from treatment in metastatic castration-resistant prostate cancer. Clin. Cancer Res. 2008;14:6302–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18829513. Accessed July 23, 2012.

3. Cristofanilli M, Hayes DF, Budd GT, Ellis MJ, Stopeck A, Reuben JM, Doyle G V, Matera J, Allard WJ, Miller MC, Fritsche H a, Hortobagyi GN, Terstappen LWMM. Circulating tumor cells: a novel prognostic factor for newly diagnosed metastatic breast cancer. J. Clin. Oncol. 2005;23:1420–30. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15735118. Accessed July 16, 2014.

4. Allard WJ, Matera J, Miller MC, Repollet M, Connelly MC, Rao C, Tibbe AGJ, Uhr JW, Terstappen LWMM. Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases. Clin. Cancer Res. 2004;10:6897–904. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15501967. Accessed July 15, 2012.

5. Cohen SJ, Punt CJ a, Iannotti N, Saidman BH, Sabbath KD, Gabrail NY, Picus J, Morse M, Mitchell E, Miller MC, Doyle G V, Tissing H, Terstappen LWMM, Meropol NJ. Relationship of circulating tumor cells to tumor response, progression-free survival, and overall survival in patients with metastatic colorectal cancer. J. Clin. Oncol. 2008;26:3213–21. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18591556. Accessed July 17, 2014.

6. Adams DL, Zhu P, Makarova O V., Martin SS, Charpentier M, Chumsri S, Li S, Amstutz P, Tang C-M. The systematic study of circulating tumor cell isolation using lithographic microfilters. RSC Adv. 2014;4:4334. Available at: http://xlink.rsc.org/?DOI=c3ra46839a. Accessed August 19, 2014.

7. Dickson MN, Tsinberg P, Tang Z, Bischoff FZ, Wilson T, Leonard EF. Efficient capture of circulating tumor cells with a novel immunocytochemical microfluidic device. Biomicrofluidics 2011;5:34119–3411915. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3364832&tool=pmcentrez&rendertype=abstract. Accessed August 19, 2014.

8. Mikolajczyk SD, Millar LS, Tsinberg P, Coutts SM, Zomorrodi M, Pham T, Bischoff FZ, Pircher TJ. Detection of EpCAM-Negative and Cytokeratin-Negative Circulating Tumor Cells in Peripheral Blood. J. Oncol. 2011;2011:252361. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3090615&tool=pmcentrez&rendertype=abstract. Accessed August 19, 2014.

9. Xu W, Cao L, Chen L, Li J, Zhang X-F, Qian H-H, Kang X-Y, Zhang Y, Liao J, Shi L-H, Yang Y-F, Wu M-C, Yin Z-F. Isolation of circulating tumor cells in patients with hepatocellular carcinoma using a novel cell separation strategy. Clin. Cancer Res. 2011;17:3783–93. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21527564. Accessed July 25, 2014.

10. Scholtens TM, Schreuder F, Ligthart ST, Swennenhuis JF, Greve J, Terstappen LWMM. Automated identification of circulating tumor cells by image cytometry. Cytometry. A 2012;81:138–48. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22170812. Accessed October 22, 2012.

11. Svensson C-M, Hübler R, Figge MT. Automated Classification of Circulating Tumor Cells and the Impact of Interobsever Variability on Classifier Training and Performance. J. Immunol. Res. 2015;2015:1–9. Available at: http://www.hindawi.com/journals/jir/2015/573165/.

12. Svensson C-M, Krusekopf S, Lücke J, Thilo Figge M. Automated detection of circulating tumor cells with naive bayesian classifiers. Cytometry. A 2014. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24733633. Accessed April 16, 2014.

13. Gorges TM, Tinhofer I, Drosch M, Röse L, Zollner TM, Krahn T, von Ahsen O. Circulating tumour cells escape from EpCAM-based detection due to epithelial-to-mesenchymal transition. BMC Cancer 2012;12:178. Available at: http://www.biomedcentral.com/1471-2407/12/178.

14. Wang Z, Li Y, Kong D, Banerjee S, Ahmad A, Azmi AS, Ali S, Abbruzzese JL, Gallick GE, Sarkar FH. Acquisition of epithelial-mesenchymal transition phenotype of gemcitabine-resistant pancreatic cancer cells is linked with activation of the notch signaling pathway. Cancer Res. 2009;69:2400–7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2657919&tool=pmcentrez&rendertype=abstract. Accessed November 14, 2014.

15. Huang C, Smith JP, Saha TN, Rhim AD, Kirby BJ. Characterization of microfluidic shear-dependent epithelial cell adhesion molecule immunocapture and enrichment of pancreatic cancer cells from blood cells with dielectrophoresis. Biomicrofluidics 2014;8:044107. Available at: http://scitation.aip.org/content/aip/journal/bmf/8/4/10.1063/1.4890466.

16. Huang C, Santana SM, Liu H, Bander NH, Hawkins BG, Kirby BJ. Characterization of a hybrid dielectrophoresis and immunocapture microfluidic system for cancer cell capture. Electrophoresis 2013:2970–2979. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23925921. Accessed November 13, 2013.

17. Huang C, Liu H, Bander NH, Kirby BJ. Enrichment of prostate cancer cells from blood cells with a hybrid dielectrophoresis and immunocapture microfluidic system. Biomed. Microdevices 2013;15:941–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23807279. Accessed January 14, 2014.

18. Smith JP, Huang C, Kirby BJ. Enhancing sensitivity and specificity in rare cell capture microdevices with dielectrophoresis. Biomicrofluidics 2015;9:1–14.

19. Morgan H, Green NG. AC Electrokinetics: Colloids and Nanoparticles. Hertfordshire, England: Research Studies Press Ltd.; 2003.

20. Kirby BJ. Micro- and nanoscale fluid mechanics : Transport in Microfluidic Devices. 1st ed. Cambridge: Cambridge University Press; 2010. Available at: http://www.kirbyresearch.com/index.cfm/wrap/textbook/microfluidicsnanofluidics.html.

21. Jones TB. Electromechanics of Particles. Cambridge: Cambridge University Press; 1995. Available at: http://dx.doi.org/10.1017/CBO9780511574498. Accessed February 18, 2014.

22. Ratanachoo K, Gascoyne PRC, Ruchirawat M. Detection of cellular responses to toxicants by dielectrophoresis. Biochim. Biophys. Acta 2002;1564:449–58. Available at:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2726261&tool=pmcentrez&rendertype=abstract.

23. Thege FI, Lannin TB, Saha TN, Tsai S, Kochman ML, Hollingsworth M a, Rhim AD, Kirby BJ. Microfluidic immunocapture of circulating pancreatic cells using parallel EpCAM and MUC1 capture: characterization, optimization and downstream analysis. Lab Chip 2014;14:1775–84. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24681997. Accessed October 4, 2014.

24. Nagrath S, Sequist L V, Maheswaran S, Bell DW, Irimia D, Ulkus L, Smith MR, Kwak EL, Digumarthy S, Muzikansky A, Ryan P, Balis UJ, Tompkins RG, Haber D a, Toner M. Isolation of rare circulating tumour cells in cancer patients by microchip technology. Nature 2007;450:1235–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3090667&tool=pmcentrez&rendertype=abstract. Accessed November 13, 2013.

25. Shah AN, Summy JM, Zhang J, Park SI, Parikh NU, Gallick GE. Development and characterization of gemcitabine-resistant pancreatic tumor cells. Ann. Surg. Oncol. 2007;14:3629–37. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17909916. Accessed November 18, 2014.

26. Mata TM, Martins A a., Caetano NS. Microalgae for biodiesel production and other applications: A review. Renew. Sustain. Energy Rev. 2010;14:217–232. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1364032109001646.

27. Chisti Y. Biodiesel from microalgae beats bioethanol. Trends Biotechnol. 2008;26:126–131. Available at: http://dx.doi.org/10.1016/j.biotechadv.2007.02.001.

28. Huang G, Chen F, Wei D, Zhang X, Chen G. Biodiesel production by microalgal biotechnology. Appl. Energy 2010;87:38–46. Available at: http://dx.doi.org/10.1016/j.apenergy.2009.06.016.

29. Pruvost J, Van Vooren G, Cogne G, Legrand J. Investigation of biomass and lipids production with Neochloris oleoabundans in photobioreactor. Bioresour. Technol. 2009;100:5988–5995. Available at: http://dx.doi.org/10.1016/j.biortech.2009.06.004.

30. Hadady H, Wong JJ, Hiibel SR, Redelman D, Geiger EJ. High frequency dielectrophoretic response of microalgae over time. Electrophoresis 2014:1–18. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25229637.

31. Michael K a., Hiibel SR, Geiger EJ. Dependence of the dielectrophoretic upper crossover frequency on the lipid content of microalgal cells. Algal Res. 2014;6:17–21. Available at: http://dx.doi.org/10.1016/j.algal.2014.08.004.

32. Bono MS, Ahner B a, Kirby BJ. Detection of algal lipid accumulation due to nitrogen limitation via dielectric spectroscopy of Chlamydomonas reinhardtii

suspensions in a coaxial transmission line sample cell. Bioresour. Technol. 2013;143:623–31. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23845710. Accessed January 21, 2014.

33. Bono MS. DIELECTRIC MEASUREMENT OF ALGAL LIPID CONTENT FOR BIODIESEL PRODUCTION. 2015.

34. Broche LM, Labeed FH, Hughes MP. Extraction of dielectric properties of multiple populations from dielectrophoretic collection spectrum data. Phys. Med. Biol. 2005;50:2267–74. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-20044380970&partnerID=tZOtx3y1.

35. Muller T, Schnelle T, Fuhr G. Dielectric single cell spectra in snow algae. Polar Biol 1998;20:303–310.

36. Wu Y, Huang C, Wang L, Miao X, Xing W, Cheng J. Electrokinetic system to determine differences of electrorotation and traveling-wave electrophoresis between autotrophic and heterotrophic algal cells. Colloids Surfaces A Physicochem. Eng. Asp. 2005;262:57–64.

CHAPTER 2

Comparison and Optimization of Machine Learning Methods for Automated Classification of Circulating Tumor Cells[1]

2.1 Abstract

Advances in rare cell capture technology have made possible the interrogation of circulating tumor cells (CTCs) captured from whole patient blood. However, locating captured cells in the device by manual counting bottlenecks data processing by being tedious (hours per sample) and compromises the results by being inconsistent and prone to user bias. Some recent work has been done to automate the cell location and classification process to address these problems, employing image processing and machine learning (ML) algorithms to locate and classify cells in fluorescent microscope images. However, the type of machine learning method used is a part of the design space that has not been thoroughly explored. Thus, we have trained four ML algorithms on three different datasets. The trained ML algorithms locate and classify thousands of possible cells in a few minutes rather than a few hours, representing an order of magnitude increase in processing speed performance. Furthermore, some algorithms have a significantly ($p<0.05$) higher area under the receiver operating characteristic curve (performance metric) than other algorithms. Optimal algorithm selection depends on the peculiarities of the individual dataset, indicating the need of a careful comparison and optimization of algorithms for individual image classification tasks.

[1] Parts of this chapter have been accepted for publication: Comparison and optimization of machine learning methods for automated classification of circulating tumor cells. TB Lannin, FI Thege, BJ Kirby. *Cytometry A*.

2.2 Introduction

2.2.1 Motivation

In 2014, over 580,000 deaths are estimated to occur due to cancer (1), and 90% of reported deaths have been from metastatic disease (2). One putative mode of metastasis is by way of circulating tumor cells (CTCs), in which cells detach from a tumor, disseminate through the bloodstream, and form new tumors at distant sites. To gain insight into this part of the metastatic process and to obtain cancer cell samples for other analyses, numerous groups have developed microfluidic tools to capture and purify CTCs from patient blood, as detailed in several reviews (3–5). These CTCs are exceedingly rare, of the order of 1-10's of CTCs per mL versus 10^6-10^7 leukocytes per mL, making the capture of pure samples a technical challenge (6). The near overwhelming number of potentially contaminating leukocytes often necessitates a post-capture analysis step in which CTCs are discriminated from captured contaminating cells and debris via fluorescent microscopy.

Simple enumeration of circulating tumor cells has prognostic value (7), but CTC capture has yet to inform clinical treatment (8). Functional assays are informative, but require high resolution images of CTCs (9) . To obtain high resolution images with a throughput high enough for a clinical setting, one must quickly discriminate CTCs from contaminants. Using a custom flow-based image cytometer, Kim et al. (10) was able to identify xenograft cancer cell clusters in a mouse model based on combined morphological and fluorescence parameters. Alternatively, Hsieh et al. (11), Tibbe et al. (12), and Scholtens et al. (13) have all developed specialized hardware for the task of locating static, captured cells on a slide or in a device via a high speed, low

resolution scan. In lieu of such specialized hardware, automated image classification algorithms, like the ones we analyze in this paper, can enable a greater breadth of researchers to locate cells on chip with conventional microscopes.

The gold standard for CTC capture and enumeration, CellSearch, (7,14–16) and other CTC isolation systems (17–21) require expert users to review images of candidate cells; however, manual classification is limited by user fatigue and inter-user variability. Automating the cell identification and classification from low resolution microscope images can make the process fast, on the order of minutes rather than hours, and perfectly repeatable. Furthermore, automation makes data organization and bookkeeping easy, and algorithms can quantify the certainty of the results.

2.2.2 State of the art

A rich literature describes machine vision for automated analysis of cells and tissues. The applications are numerous. A few examples include the classification of blood cells via neural networks (22), grading of cervical intraepithelial neoplasia by extracting geometrical features (23), real time monitoring of yeast cell density and viability via support vector machines (24), monitoring hematopoietic stem cells in image time sequences via dynamic background compensation (25), localization of sub-cellular components via threshold adjacency statistics (26), and observation of textural differences between cancer cells and blood cells (27,28).

Most automated cell classification algorithms work on a 3-step process. First, objects that are cell-sized and cell-shaped are segmented from or detected against the background. Segmentation is often accomplished by setting an Otsu intensity threshold

to define bright regions against a dark background (29). Second, a vector of features is extracted from the event. In the simplest case, these features could be the intensity value of each pixel in an image patch centered on the event. Alternatively, these features could be quantitative values that parallel the information that humans extract when they discriminate between different objects, for example, the area of the segmented region or the intensities of different fluorescent stains within the segmented region. Third, a machine learning algorithm, trained by a set of manual classifications of these objects, determines a rule that emulates a set of reference human classifications as accurately as possible given the information contained in the features.

Toolkits, detailed in the review paper of Shamir et al. (30), have been developed to make cell segmentation and automated identification available to wider audience. Of note, CellProfiler (31–33) operates on a set of extracted image features (e.g. object area, intensity, correlation between different colors) , and it makes use of GentleBoosted regression stumps to classify cells. Also, the creators incorporate a graphical user interface friendly to non-programmers. Wndchrm extracts a different set of features, mostly histograms or coefficients of various image transforms (34). Wndchrm then uses a Fischer information-weighted nearest neighbors algorithm in the feature space to classify new images. CellExplorer (35) makes use of support vector machines to identify nuclei in 3D confocal image stacks. Ozkumur et al. (36) and Wang et al. (37) employ proprietary software (BioView Ltd., Rehovot, Israel, and Definiens Cognition Network Technology, Munich, Germany, respectively) in their analyses of cancer cell images.

The publicly available toolkits have not been directly applied to the problem of automated identification of circulating tumor cells, but very similar techniques have been employed in the literature. Scholtens et al. present (and Ligthart et al. validate) an image cytometer for the identification of circulating tumor cells. It extracts similar features to those of CellProfiler, with a random forest (RF) classifier for the automated classification of CTCs purified by the CellSearch system (13,38–40). Also, Svensson et al. explore a semisupervised Bayesian Classifier to a support vector machine, both operating on intensity histograms (one of the feature components of Wndchrm) to classify images of cancer cells captured on a functionalized and structured medical wire (41).

In these works, substantial attention has been given to the image processing techniques for extracting relevant features, but less attention has been paid to the choice of machine learning algorithm used in the classification step. In particular, Scholtens et al. justify their use of a random forest classifier by the algorithm's prevalence in the literature and its comparable performance to that of multiple users compared to each other, but make no further exploration of the selection of machine learning method as an element of the design space (38). Only Svensson et al. directly compared different machine learning algorithms, a support vector machine and a Bayesian classifier; however, their dataset was rather small (617 events used in machine learning, split 75%-25% training-testing), and the only statistically significant difference in accuracy they observed occurred when they artificially lowered the percentage of labeled data to 5%.

As will be demonstrated in this work, exploration and thorough optimization of different machine learning algorithms becomes increasingly important when the data is noisy, feature nonlinearities exist, or disagreements occur within training data from multiple users. For example, rapidly imaging cells in the 3D, complicated geometry of novel microfluidic devices such as our geometrically-enhanced differential immunocapture device (42), or needing to compromise staining for cell identification with staining for clinically impactful downstream analyses (9) results in many ambiguous images of captured cells (Figure 2.1).

2.2.3 Aim of this work

Our aims for this work were threefold. First, we compared four machine learning methods to more thoroughly explore that element of the design space. We investigated: 1) a random forest classifier as Scholtens et al. and Svensson et al. (2015) have done (39,43), and similar to the GentleBoosted regression stumps as CellProfiler has done (32–34), 2) a K-nearest neighbors algorithm, similar to the classification step of Wndchrm (35), 3) a Bayesian classifier, similar to the classification step of Svensson et al. (42), and a support vector machine, as is employed in the classification step of CellExplorer and the baseline algorithm of Svensson et al. (36,42). Second, we determined whether extracting a set of image features (as done by Scholtens et al.) is superior to using a more raw set of features: a list of the intensity of each pixel of an image patch (as done by Svensson et al. 2015). These comparisons were performed on three datasets: 1) a model dataset of cancer cell lines spiked into isolated white blood cells to serve as a baseline with two distinct, unambiguous cell populations, 2) a set of

cancer cells captured on chip from blood samples from a pancreatic cancer patient, and 3) a set of cancer cells captured on chip from blood samples from prostate cancer patients. The varying degrees of ambiguity (as reflected by the fraction of disputed examples) within these three datasets provided a platform to observe how performances of the algorithms degrade in comparison to manual classifications. Third, we explored how robustly one may apply a trained machine learning algorithm out of context by training the algorithm on one dataset, but testing it on another set.

2.3 Materials and Methods

2.3.1 Experiment overview

Three sets of data were imaged. One set of data consisted of LNCaP model prostate cancer cells spiked into isolated peripheral mononuclear blood cells from a healthy donor and imaged on a coverslip. The second and third datasets are collected from human cancer patients, stained, and imaged on the GEDI device. The second set is three pancreatic cancer patient samples, detailed in another work (43). The third set consists of three blood samples obtained from metastatic prostate cancer patients. All samples were obtained with informed consent from patients in accordance with IRB-approved protocols. We will refer to the LNCaP/PBMC set as set A, the pancreatic cancer patient set as set B, and the prostate cancer patient set as set C.

For sets B and C, blood samples were run through geometrically-enhanced differential immunocapture (GEDI) devices described in Gleghorn et al., Thege et al., and Kirby et al. (9,44,45). The GEDI capture antibody was a combination of a monoclonal antibody for human epithelial cell adhesion molecule (Santa Cruz

Biotechnology, Santa Cruz) and/or a monoclonal antibody for cancer-specific mucin, MUC1 (provided by Michael A. Hollingsworth, University of Nebraska) for the set B and a monoclonal antibody for J591 prostate specific membrane antigen (provided by Dr. Neil Bander, Weill Cornell Medical College) for set C.

2.3.2 Sample and specimen description

For set A, LNCaP prostate cancer cells were acquired from ATCC and cultured according to provider specifications. These cells were spiked into a batch of peripheral mononuclear blood cells (PBMCs) isolated from healthy donors. The LNCaPs serve as model cancer cells and the PBMCs serve as model blood cells for an unambiguous control with two distinct populations of cells. The two types of cells were spiked onto a slide and were fixed in PHEMO fixative then blocked with normal goat serum. The fixed cells were stained for CD45 with a primary (BD Biosciences) and an AlexaFluor (AF) 488 conjugated secondary antibody (Life Technologies), for pan-cytokeratin with a primary (Biolegend, San Diego, CA) and a CF594 Mix-N-Stain conjugation kit (Biotium, Inc., Hayward, CA) and for nuclei using DAPI (Life Technologies).

For set B, the covers of the GEDI devices were removed, and the samples were fixed in 2% PFA in 50% PHEM buffer (60 mM PIPES, 25 mM HEPES, 10 mM EGTA and 2 mM MgCl2) for 15 minutes and blocked in 6% BSA and 10% normal goat serum in PBS for 1 hour. After staining of surface markers, the samples were permeabilized with 0.25% (w/w) Triton x-100. The samples were stained with, for MUC4, a primary (ab60720; abcam, Cambridge, MA, USA) and an AlexaFluor 488 conjugated secondary antibody (Invitrogen), for CD45, a Qdot-800 conjugated

antibody (Invitrogen), for CK, a CF543 (Biotium, Hayward, CA, USA) conjugated anti-pan-CK antibody (C11; BioLegend, San Diego, CA, USA) and, for DNA/nuclei, DAPI (Invitrogen).

For the set C, the covers of the GEDI devices were removed, and samples were fixed in PHEMO fixative then blocked with normal goat serum. The fixed cells were stained for CD45 with a primary (BD Biosciences) and a Qdot 800-conjugated secondary antibody (Life Technologies), for pan-cytokeratin with a primary (Biolegend, San Diego, CA) and a CF594 Mix-N-Stain conjugation kit (Biotium, Inc., Hayward, CA) and for nuclei using DAPI (Life Technologies).

2.3.3 Instrument Details

For sets A and C, a Zeiss Observer Z1 microscope stand equipped with a Yokogawa CSU-X1 spinning disk head operated with Axiovision 4.8.2 software imaged the whole slide area or capture volume of the chip with a 10x/0.3NA objective via a tiling macro. Each flurophor was imaged independently by excitation with the laser closest to the excitation peak (405nm: DAPI and Qdot 800, 488nm: AF488, 561nm: CF594) and the emission was collected through the proper filter (DAPI: BP 450/50, AF488: 525/50, CF 594: BP 629/62, and Qdot 800: LP 700).

For set B, a Zeiss LSM 5 Live Confocal Microscope operated with AIM software imaged the whole capture volume of the chip with a 10x/0.3NA objective via a tiling macro. Each flurophor was imaged independently by excitation with closest on-peak laser (405nm: DAPI, 488nm: AF 488, or 532nm: CF543) and the emission was

collected through the proper filter (DAPI: BP415-480, AF 488: BP 500-525, CF543: BP 550-615, and AF 680: LP 650).

2.3.4 Data analysis details

Pre-Screening Algorithm. We desired a pre-screening algorithm that did not explicitly segment cells based on a global threshold. Thus, we developed a spatial filtering and peak-detection MATLAB (MathWorks, Natick, MA) script that detects the center of cell-sized bright regions. First, the script normalizes each image by subtracting the median and dividing by the median absolute deviation (robust statistics analogs of mean and standard deviation) of the pixel intensities in the image. Then, it takes a maximum intensity projection in Z of the normalized DAPI images. The projected image is then filtered with a difference of Gaussians spatial bandpass filter where the bandpass parameters are the expected range of cell radii. All local maxima are then detected in this filtered image. The difference of Gaussians parameters were properly chosen such that there is sufficient smoothing to suppress the graininess (multiple local maxima from noise) within a single cell, yet not so much smoothing that adjacent cells will be blurred into a single local maximum. Across an entire image, most of these local maxima would be from random noise in the background, and the local maxima associated with CTCs will only correspond to local maxima at which both the DAPI signal and the cytokeratin (CK) or other positive identifier signal are unusually bright. Thus, the script retains only the local maxima that are higher than a threshold in both DAPI signal and CK signal, where this threshold scales with the width of the distribution of maxima from background:

$$Threshold = Median(I) + \varphi\, MAD(I) \quad (1)$$

where I is the list of pixel intensities, MAD is the median absolute deviation operator, and φ is a hand-tuned factor which allows the user to define at what probability threshold an event is unusually bright, constituting a potential CTC. (We selected $\varphi = 6$, which, for the signal-to-background of our images, was conservative enough such that no CTCs were missed during observation.) The locations of sufficiently bright local maxima serve as a list of the approximate centers of all DAPI+/CK+, cell-sized events. For each z-slice and for each channel, cut out a small (39 μm x 39μm) image patch of the unfiltered (but still normalized) image, centered on the local maximum. For training the machine learning algorithm, this image patch is presented to the user for manual classification.

General Approach of Machine Learning. Supervised machine learning for classification is a process whereby the relationship between input features and the desired output class is modeled by an algorithm. For each machine learning algorithm used in this study, two possible sets of input features were used. The first option for input features is the intensity level of each pixel in the image patch. The second option for input features is the set of features extracted by Scholtens et al. Features for each stain in the patch included the area, perimeter, eccentricity, and major axis length of the largest Otsu-thresholded region in the patch, the total intensity, max intensity, and intensity standard deviation in this thresholded region, the contrast mean, correlation, and homogeneity of the gray level co-occurrence matrix of the patch, and the entropy of the of the patch. Features for each pairwise combination of stains of the patch included the total intensity ratio, the slope of the line correlating intensities, and the R^2

value of this fit line. Extracting this set of features presents the benefit of potentially transforming the feature space in a way that provides a more robust separation between CTCs and non-CTCs at the cost discarding some of the original information in the image patch.

Using manual training examples, some parameters of the machine learning function may be internally optimized to best reproduce the known output. Additional free parameters for each method may be optimized via cycling through combinations of training on a subset of the data and validating on the remainder (Figure 2.2, parameter ranges in supplemental info). After an algorithm is thoroughly optimized, to estimate how well it may perform on new data, one may reserve a random set of the total available data for testing (e.g. 25% of the total dataset for our work).

Supervised machine learning algorithms need a ground truth. For our experiments, it was generated from a majority of votes from three manual classifications of trained experts labeling each image patch. Although more sophisticated algorithms exist that achieve performance improvements when the algorithms account for peculiarities of individual experts, such improvements have been observed only when the number of experts was large (46).

Performance Metric. Because a large fraction of examples are negative (non-CTC), performance on the test set should not be judged on accuracy, because an algorithm could achieve high accuracy (e.g. 97.5% for set C) by always predicting the negative class. That algorithm would be trivial, so it is more useful to frame performance in terms of tradeoffs between increasing the true positive rate (TPR) at the cost of increasing the false positive rate (FPR). The receiver operating

characteristic (ROC) curve maps out the achievable TPR for a specified FPR of a machine learning algorithm, and the area under the ROC curve is a performance metric (ranging from 0 to 1) by which different machine learning methods may be compared (Figure 2.3, right column). An empirical ROC with a finite data set may be constructed by sorting the examples by the output score of the machine learning algorithm, and considering all possible thresholds for this score. This process results in several discrete points along the AUC. A composite classifier with performance on a line segment between any two points on the ROC may be constructed by randomly selecting either the classifiers corresponding to the segment's endpoints (47). Thus, AUC may be computed by trapezoidal integration of the empirical ROC points.

In cases where the number of negative examples is large, as is the case for our classification task, even a moderate FPR will cause the false positives to overwhelm the true positives. For such problems, the AUC of the entire ROC is not an appropriate performance metric. Thus, we defined a performance metric to be the AUC of the ROC up to a FPR of 0.05 (Figure 2.3 left column). In this way, we optimized performance only for the part of the ROC that important to the application.

Machine Learning Algorithm Overview. We used four machine learning algorithms in the MATLAB Statistics and Machine Learning Toolbox (The MathWorks, Inc., Natick, MA): i) K-nearest neighbors, ii) support vector machine, iii) random forest classifier, and iv) Bayesian classifier of Gaussian mixture models with pre-processing by principal components analysis. Details of the operating principles of these algorithms appear in the supplemental information of this article and in other informative articles (48–53).

Briefly, K-nearest neighbors (KNN) computes a distance metric as differences between the features of the image to be classified and each point in the training dataset. Some number (K) of nearest neighbors all vote on what the class of this new image should be. K-nearest neighbors is often effective because images of the same class (CTC or non-CTC) are likely to exist in distinct regions of the feature space, defining, by proximity to training examples, regions of space that belong to CTCs or to non-CTCs.

Support vector machines (SVMs) use an optimization process to find a set of weights for a shortened list of training examples. These weights define a boundary between regions of CTCs and non-CTCs, so rather than comparing some new image to be classified to all of the training data, it only needs to be compared to this boundary. Although SVMs were originally constrained to linear boundaries, they were extended to the nonlinear regime through the use of radial basis kernel functions, which we have used in this work.

A random forest (RF) classifier builds a set of decision trees from bootstrapped samples of the training data. Bootstrapping is random resampling of the training data with replacement. Because each tree observes a different version of the training data, the procedure adds diversity to the random forest. Each decision tree is a series of yes/no splits on a different feature, breaking the feature space into successively more pure sub groups. Eventually, this process terminates, with a full set of rules for which new examples may be classified. The RF classifier aggregates the votes of each of its component trees, giving not only an estimate of how this new example should be classified, but also an estimate of the algorithm's certainty of the guess.

Our Bayesian classifier abstracts the training dataset as two probability distributions: one for CTCs and one for non-CTCs. Due to the heterogeneous nature of CTCs and non-CTCs, these probability distributions are not normal; however, they each can be well approximated by a mixture of multivariate Gaussian distributions. Although these multivariate Gaussians could be defined in the dimensionality of the original feature space, the novelty of our algorithm comes from projecting the data into a lower dimensional space via principal components to improve the computational speed, numerical stability, and classification performance of the algorithm.

All four of these algorithms have tunable parameters that control tradeoffs in their classification performance. One such tradeoff is the balance between suppressing noise and occasional misclassifications in the training data without suppressing actual subtle trends that exist in the data (Figure 2.2). In KNN, for example, if K is small, few neighbors determine the class of a new example, subjecting it to the whims of any single example. In the limit of a large K, votes are included from very distant (and potentially unrelated) neighbors, compromising performance. Both limits of K are undesirable, suggesting an optimal K that is specific to the dataset.

Statistical Analysis. If the area under the entire ROC were to be used as a performance metric, a Z-test could be used to formulate a hypothesis test against the null hypothesis of equality of the two AUCs. In lieu of an analytical method for hypothesis-testing differences in AUC up to a FPR of 0.05, bootstraps of the test data were used to compute an empirical distribution of the differences in this performance metric for pairs of algorithms via a bias-corrected and accelerated algorithm (54). Finally, each test was Bonferroni-corrected to reduce inflation of type I error in

making multiple pairwise comparisons. Bootstrapping was also used to construct the confidence intervals in Figures 2.4 and 2.5.

In summary, four machine algorithms, each with different free parameters, were optimized by 3-fold cross validation (support vector machine, Bayesian classifier), leave-one-out validation (K-nearest neighbors), or out-of-bag error (random forest). Each algorithm was trained with the aggregated classifications of all three users, training it to learn the consensus of the three users.

2.4 Results

The pre-screening algorithm detected events from the images in each of the three datasets. Then, all of the events in each dataset were manually classified three times from among four trained experts. These data were then randomly split 75% for training and 25% for testing (Table 2.1). Set A showed good agreement among users with 88% of the positive training events a unanimous vote among the experts. Set B showed modest agreement among users with 40% of positive events a unanimous vote among the experts. Set C showed poor agreement among users with only 23% of positive events a unanimous vote among the experts.

The same selection of training and test sets were given to each algorithm. Performance of the algorithms was measured by computing the AUC up to a FPR of 0.05. The ROCs of these algorithms are shown in Figure 2.3 to show the TPR/FPR tradeoff. Figure 2.4 shows the AUCs up to a FPR of 0.05 for each algorithm and notes significant differences.

The trend of good agreement on set A, moderate agreement on set B, and poor agreement on set C was paralleled by performance of automated classifiers, with the score (AUC up to FPR of 0.05) of the best classifier on set A of 0.048, the best classifier on set B of 0.024, and the best classifier on set C of 0.021.

Although not all of the comparisons among all algorithms were significant at $p<0.05$, random forest classifiers tended to perform the best. There was no clear trend regarding the performance of algorithms operating on either pixel intensities or the extracted features of Scholtens et al., suggesting that this design choice is data set-specific.

When the best algorithm on set A was tested on the images from set C (and vice versa), the out-of-context classifier significantly underperformed its properly-trained counterpart ($p<0.05$).

2.5 Discussion

Given the challenging imaging environment of the 3D GEDI device with staining protocols optimized for downstream analyses rather than simple enumeration and identification and hardware not specialized for the very specific task of scanning chips, it is worth thoroughly exploring the type of machine learning method used in order to improve the performance of automated cell classifiers. Automated cell identification algorithms are composed of three parts: 1) an initial scan to detect events as possible cells, 2) a method to convert the image of the detected event into a set of quantitative features, which are passed to 3) a machine learning algorithm which, based on a database of user-labeled events, classifies the event. Each researcher in the literature has used different sets of extracted features and different machine learning algorithms, and each group has used their system on a different dataset of images, making direct comparisons between different systems difficult.

This systematic comparison of machine learning methods appears to become particularly important when the ambiguity of the dataset increases (Table 2.1 and Figures 2.3 and 2.4). High AUC performance of all algorithms on the unambiguous cell line data is consistent with the observations of Svensson et al. (42), in which both a SVM and a Bayesian Classifier could recapitulate manual classifications with extremely high precision and recall. When all of the performances are very good, one would need a very large dataset to start observing differences, but one may not care about miniscule increases in performance when the algorithm works sufficiently well. For sets B and C, in which there exists more disagreement between manual users, the selection of algorithm becomes important. This result indicates the potential danger of

not exploring the machine learning algorithm design when the training data is ambiguous.

In this analysis, we also investigated two types of feature extraction: 1) the intensities of each pixel in an image patch centered on the event and 2) the same set of features that Scholtens et al. use in (39). In essence, the feature extraction process of Scholtens et al. is a nonlinear transform of the image patch which has a cost of losing some information by lowering the dimensionality of the data, but it has the benefit of enhancing separation between CTCs and non-CTCs in the transformed feature space. This balance between enhancing separation in between the two classes at the cost of discarding information is important for the performance of the machine learning methods. There is no general rule for striking such a balance because it depends on the type and amount of training data, so researchers must optimize their methods *ad hoc*.

We observe that the performance of an algorithm trained on unrepresentative data can drop significantly when compared to the algorithm applied to its own test set. To avoid this risk, one must retrain data when the type of cells under observation or the way in which the cells were imaged changes; consistent staining and imaging experimental conditions are necessary to be certain of the valid operation of the algorithm. In particular, the use of cell lines spiked into isolated white blood cells and imaged on a cover slip alone may not serve as a good training model for cancer cells obtained from patient samples and imaged in a microfluidic device.

Future work could include testing other types of extracted features. Wndchrm makes use of a set of moments, Zernike features, and multi-scale histograms, which are likely to be invariant descriptors of cells (35). Svensson et al. employ intensity

histograms, which they have shown to be effective (42). The trouble with testing many feature sets with many algorithms is the number of pairwise statistical comparisons that must be made grows in a factorial manner, necessitating a huge dataset to draw significant conclusions.

For more future work, Svensson et al. and Jones et al. both begin to address an important point in the automated cell classification problem; the new image data is continually being obtained could improve the performance of algorithms, but manually classifying all of this new unlabeled data would be tedious (42,55). Svensson et al. incorporate the use of unlabeled data in addition to previously labeled data to enhance the estimates of the probability distributions in their Bayesian classifier. Jones et al. propose an iterative machine learning process whereby a user classifies some of the image data, the algorithm proposes a tentative rule, and then the user refines the rule iteratively with new batches of images. Because the algorithm returns batches of borderline examples, it spares the user from tediously classifying obvious examples that would not provide the algorithm with much more information. Machine learning on a partially labeled dataset or through an iterative process is not specific to the methods used by these researchers, and can be done with other methods, for example, a transductive SVM (56) trained iteratively. These iterative feedback algorithms would still, nonetheless, benefit from the systematic and rigorous comparison and thorough optimization discussed here.

2.6 Acknowledgements

The authors would like to acknowledge former classmates Thomas Schmidt, Max Spector, and Yo Han Ko as well as Dr. Thorsten Joachims and Dr. Kilian Weinberger their discussions of machine learning algorithms. The authors would also like to acknowledge colleagues Erica D. Pratt, Matthew S. Sung, Paraskevi Giannakakou, Guang Yu Lee, David Nanus, and Andrew Rhim for their assistance in obtaining and/or classifying data.

2.7 Figures, Tables, and their Legends

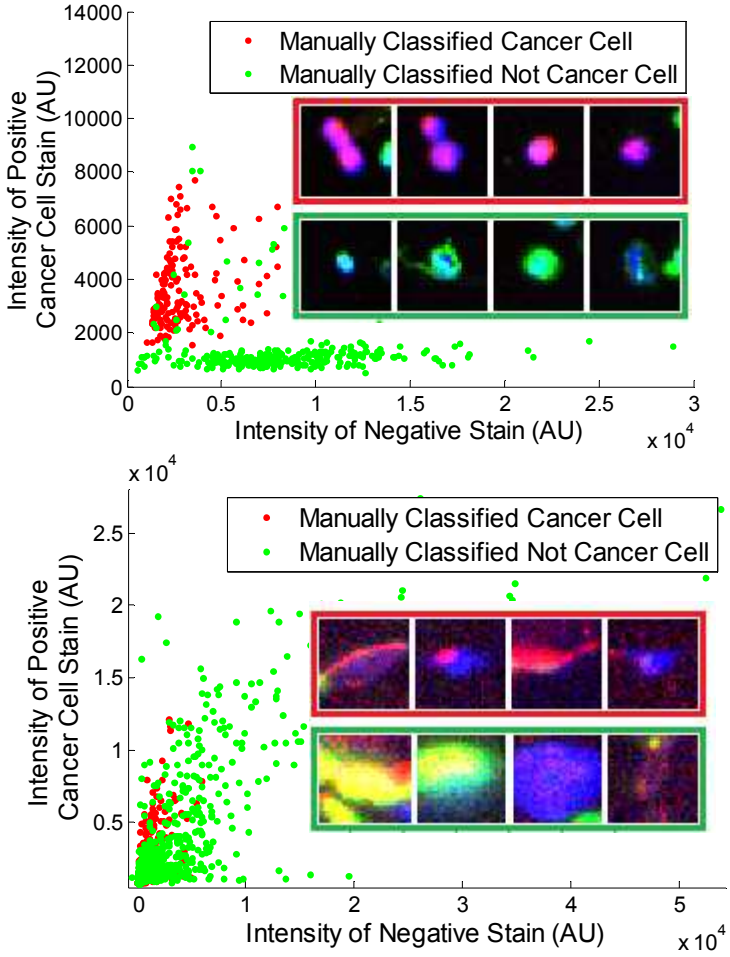

Figure 2.1 – Scatterplots of maximum intensities of positive stain vs. negative stain for two datasets. Top: model prostate cancer cell line (LNCaPs) and peripheral mononuclear blood cells spiked onto a coverslip (dataset A in text). Bottom: pancreatic cancer patient cells, and blood cells/debris captured in a GEDI device (dataset B in the text). Note that for the high signal-to-noise model system, cancer cells and non-cancer cells exist as two nearly separate populations, whereas in the moderate signal-to-noise real system, there is far more overlap in the populations. Insets show randomly selected image patches of cancer cells (outlined in red) and non-cancer cell events (outlined in green). Blue: DAPI, Red: Cytokeratin, Green: CD45. Patch width = 39 μm.

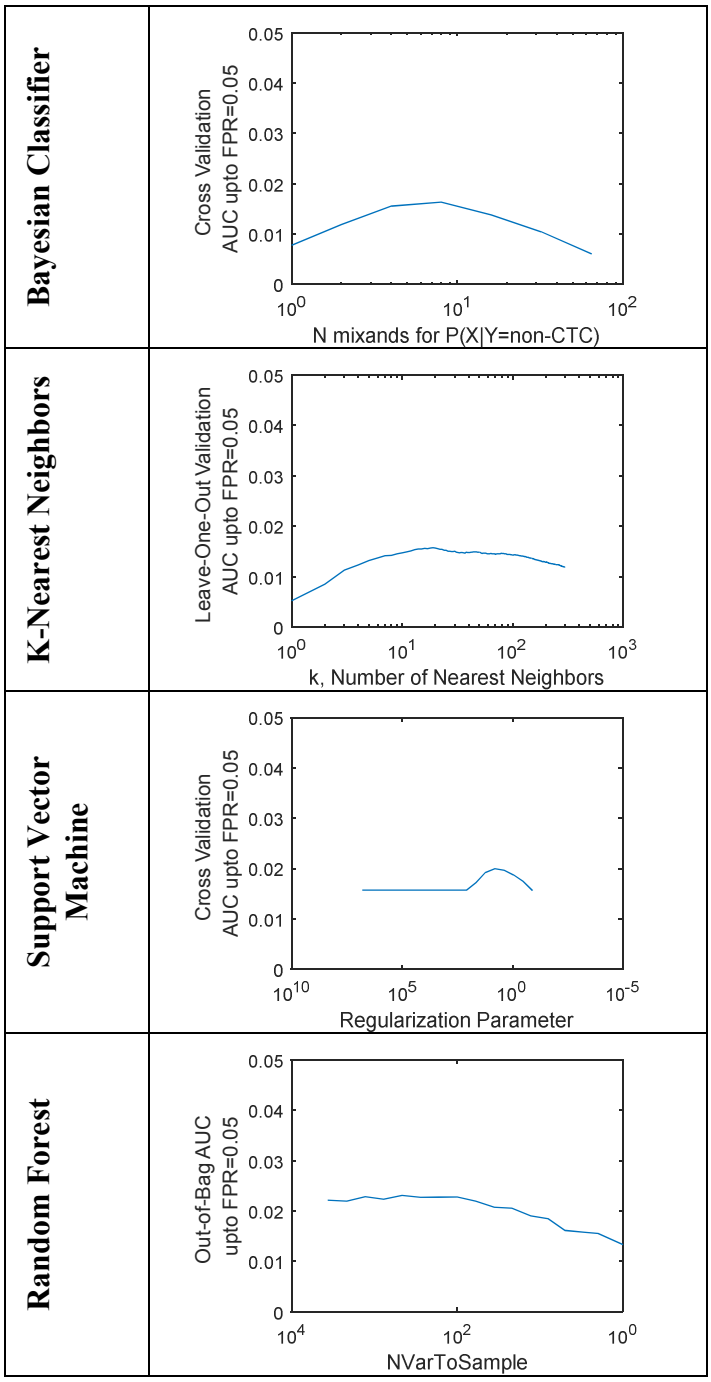

Figure 2.2 – The effect of varying algorithm-specific regularization parameters on the validation performance of algorithms (as measured by area under the receiver operating characteristic curve up to a false positive rate of 0.05) on data set B (Pancreatic Cancer Patient data). Moving to the right on these x-axes would reduce noise at the cost of averaging over subtle trends. For these plots, all other tunable parameters are held constant at their global optimal values.

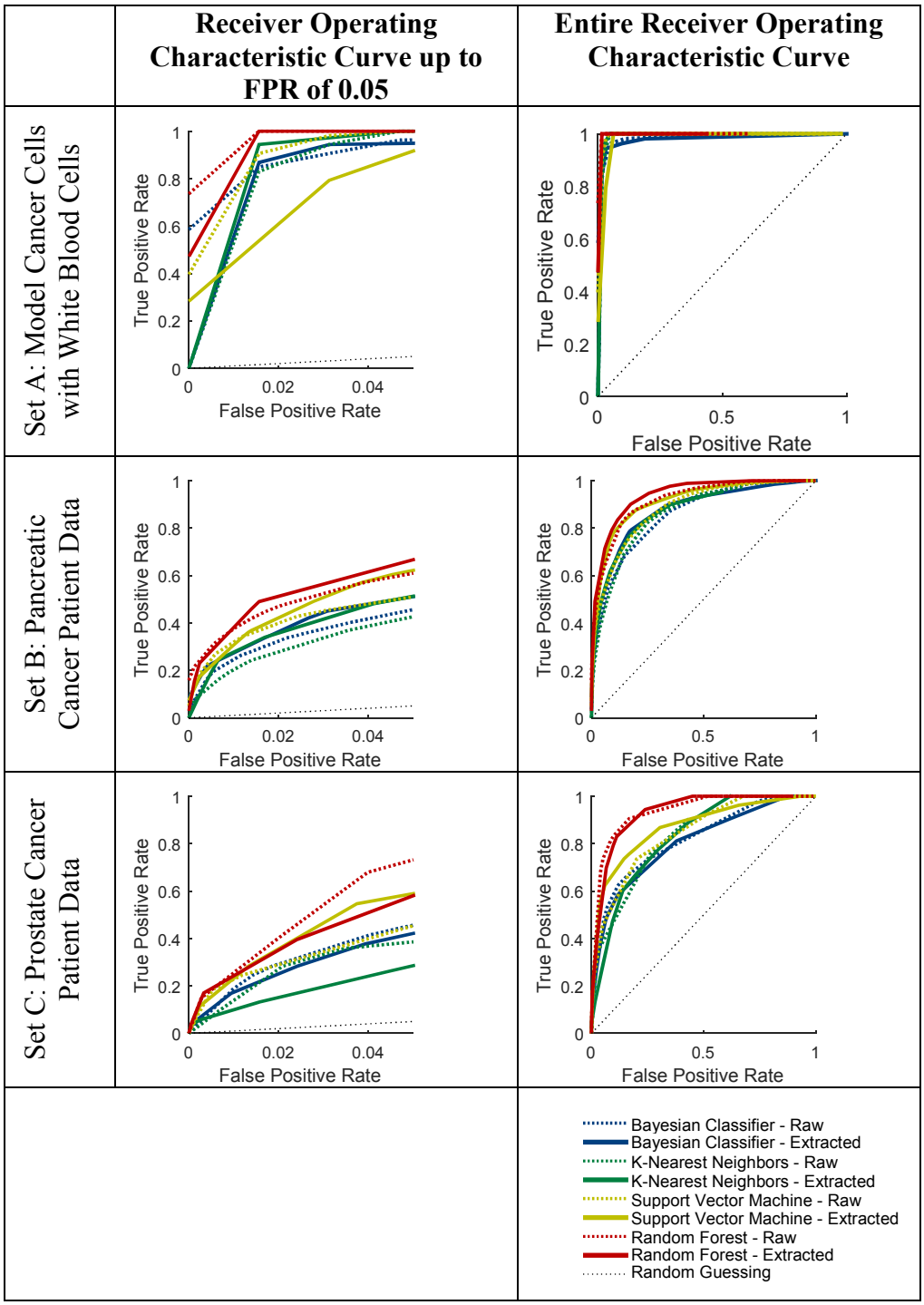

Figure 2.3 – From three different test sets, receiver operating characteristic (ROC) curves of four machine learning algorithms (Bayesian Classifier, K-Nearest Neighbors, Support vector machine, and Random Forest) each operating on either raw pixel intensities or features extracted by image processing. Also shown is the line TPR = FPR, which corresponds to random guessing. Left: The portion of the ROC up to a false positive rate of 0.05, the area under which served as the performance metric for these algorithms. Right: The entire ROC for each of the algorithms.

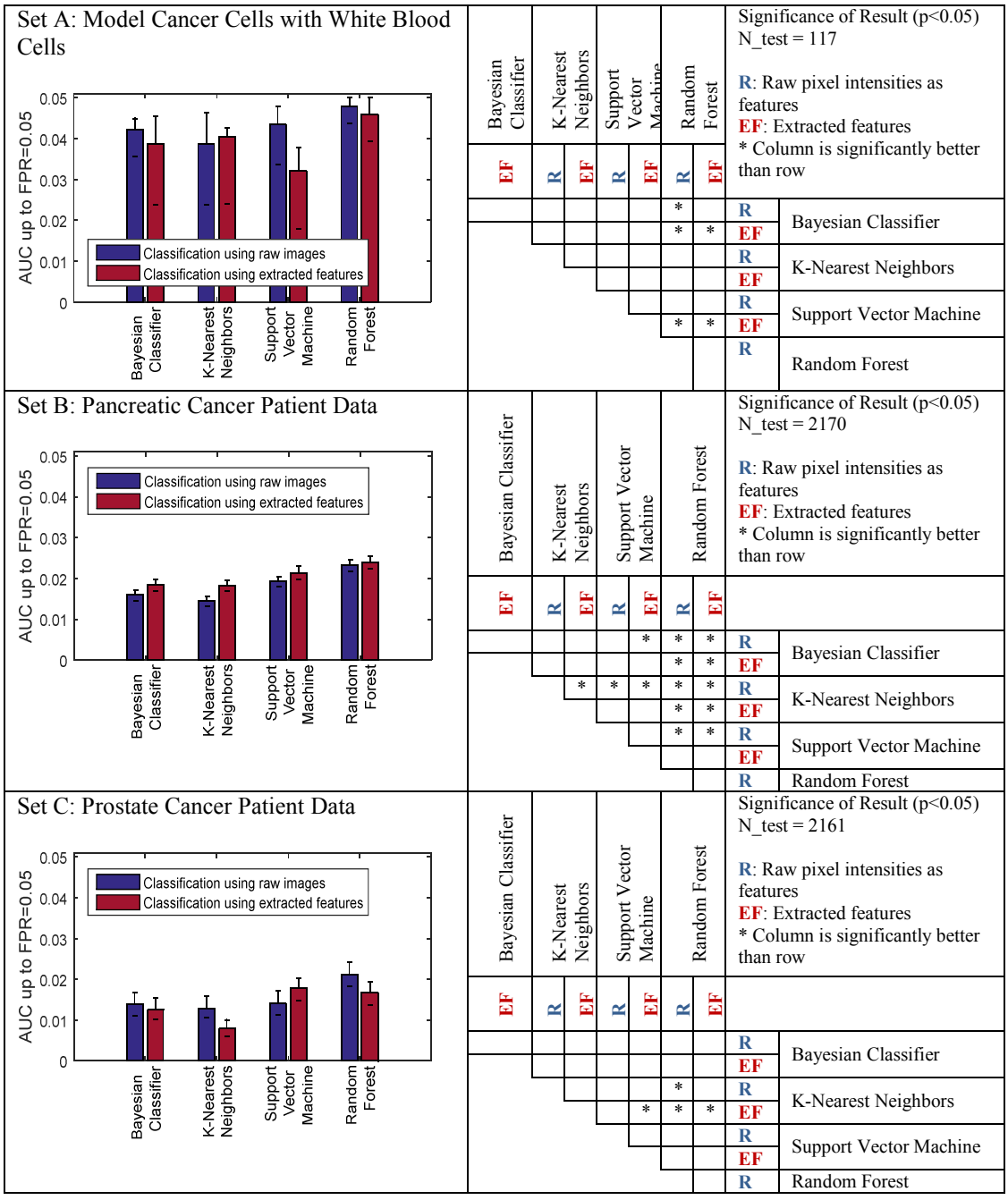

Figure 2.4 – Left: Area under the receiver operating characteristic curve (AUC) up to a false positive rate (FPR) of 0.05 for four machine learning algorithm classifications on whether image patches contain cultured cancer cell line cells (set A), a pancreatic cancer patient cells (set B), or prostate cancer patient cell (set C). Error bars indicate bootstrapped 68.2% confidence intervals. Right: the Bonferroni-corrected significance ($p<0.05$) of all pairwise comparisons of techniques.

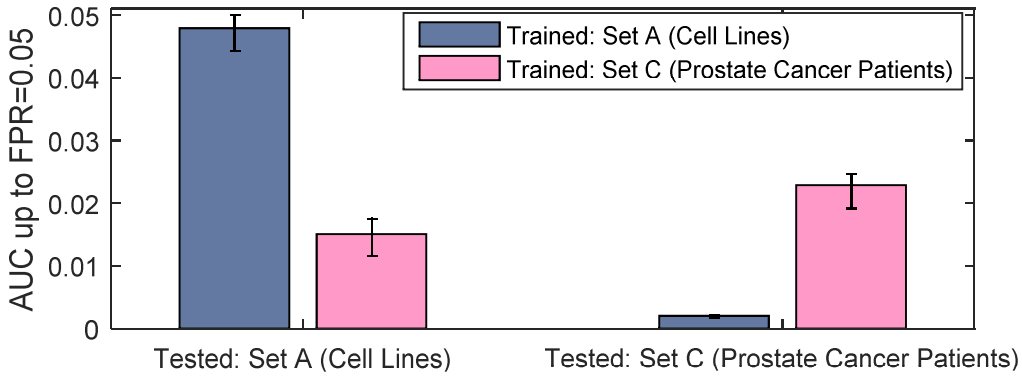

Figure 2.5 – Performance (as measured by area under the receiver operating characteristic curve (AUC) up to a false positive rate (FPR) of 0.05) of the random forest classifier operating on pixel intensities as features. The classifiers from sets A and C were tested on the test sets from its own and a different dataset. Error bars are bootstrapped confidence intervals (68.2%). Note that set B was not able to be compared because two positive identifier stains were used (for cytokeratin and for MUC1) rather than one stain (cytokeratin) stain for sets A and C. Thus, the dimensionality of set B does not match the other datasets. Both comparisons are significant with $p<0.05$.

Table 2.1 – Description of sizes, manual classifications, and training/testing allocations of the datasets used to analyze machine learning algorithms. Positive events are defined as a majority vote among the three trained experts. Disagreements are the number of events in which one user had a different label from the other two.

Dataset	Set A: Cell Lines	Set B: Pancreatic Cancer Patient CTCs	Set C: Prostate Cancer Patient CTCs
Number of Training Events	352	6509	6484
Number of Positive Training Events	125	1543	148
Number of Disputed Positive Training Events	15	917	113
Number of Disputed Negative Training Events	7	2260	364
Number of Testing Events	117	2170	2161
Number of Positive Testing Events	53	512	53
Number of Disputed Positive Testing Events	7	310	40
Number of Disputed Negative Testing Events	0	773	122

2.8 References

1. American Cancer Society. Cancer Facts & Figures.; 2014. Available at: http://www.cancer.org/Research/CancerFactsStatistics/CancerFactsFigures2014/cancer-facts-and-figures-2014.pdf.

2. Mehlen P, Puisieux A. Metastasis: a question of life or death. Nat. Rev. Cancer 2006;6:449–58. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16723991. Accessed August 15, 2013.

3. Pratt ED, Huang C, Hawkins BG, Gleghorn JP, Kirby BJ. Rare Cell Capture in Microfluidic Devices. Chem. Eng. Sci. 2011;66:1508–1522. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3082151&tool=pmcentrez&rendertype=abstract. Accessed October 4, 2012.

4. Li P, Stratton ZS, Dao M, Ritz J, Huang TJ. Probing circulating tumor cells in microfluidics. Lab Chip 2013;13:602–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23306378. Accessed August 6, 2013.

5. Zhang Z, Nagrath S. Microfluidics and cancer: are we there yet? Biomed. Microdevices 2013;15:595–609. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23358873. Accessed May 14, 2013.

6. Den Toonder J. Circulating tumor cells: the Grand Challenge. Lab Chip 2011;11:375–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21206959. Accessed March 26, 2013.

7. De Bono JS, Scher HI, Montgomery RB, Parker C, Miller MC, Tissing H, Doyle G V, Terstappen LWWM, Pienta KJ, Raghavan D. Circulating tumor cells predict survival benefit from treatment in metastatic castration-resistant prostate cancer. Clin. Cancer Res. 2008;14:6302–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18829513. Accessed July 23, 2012.

8. Parkinson DR, Dracopoli N, Petty BG, Compton C, Cristofanilli M, Deisseroth A, Hayes DF, Kapke G, Kumar P, Lee JS, Liu MC, McCormack R, Mikulski S, Nagahara L, Pantel K, Pearson-White S, Punnoose E a, Roadcap LT, Schade AE, Scher HI, Sigman CC, Kelloff GJ. Considerations in the development of circulating tumor cell technology for clinical use. J. Transl. Med. 2012;10:138. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3478228&tool=pmcentrez&rendertype=abstract. Accessed June 26, 2013.

9. Kirby BJ, Jodari M, Loftus MS, Gakhar G, Pratt ED, Chanel-Vos C, Gleghorn JP, Santana SM, Liu H, Smith JP, Navarro VN, Tagawa ST, Bander NH, Nanus DM, Giannakakou P. Functional characterization of circulating tumor cells with a prostate-cancer-specific microfluidic device. PLoS One 2012;7:e35976. Available at:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3338784&tool=pmcentrez&rendertype=abstract. Accessed November 14, 2012.

10. Kim H, Terazono H, Nakamura Y, Sakai K, Hattori A, Odaka M, Girault M, Arao T, Nishio K, Miyagi Y, Yasuda K. Development of on-chip multi-imaging flow cytometry for identification of imaging biomarkers of clustered circulating tumor cells. PLoS One 2014;9:e104372. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4139271&tool=pmcentrez&rendertype=abstract. Accessed October 7, 2014.

11. Hsieh H Ben, Marrinucci D, Bethel K, Curry DN, Humphrey M, Krivacic RT, Kroener J, Kroener L, Ladanyi A, Lazarus N, Kuhn P, Bruce RH, Nieva J. High speed detection of circulating tumor cells. Biosens. Bioelectron. 2006;21:1893–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16464570. Accessed March 13, 2013.

12. Tibbe a G, de Grooth BG, Greve J, Liberti P a, Dolan GJ, Terstappen LW. Optical tracking and detection of immunomagnetically selected and aligned cells. Nat. Biotechnol. 1999;17:1210–3. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10585720.

13. Scholtens TM, Schreuder F, Ligthart ST, Swennenhuis JF, Tibbe AGJ, Greve J, Terstappen LWMM. CellTracks TDI: an image cytometer for cell characterization. Cytometry. A 2011;79:203–13. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21337699. Accessed November 6, 2012.

14. Cristofanilli M, Hayes DF, Budd GT, Ellis MJ, Stopeck A, Reuben JM, Doyle G V, Matera J, Allard WJ, Miller MC, Fritsche H a, Hortobagyi GN, Terstappen LWMM. Circulating tumor cells: a novel prognostic factor for newly diagnosed metastatic breast cancer. J. Clin. Oncol. 2005;23:1420–30. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15735118. Accessed July 16, 2014.

15. Allard WJ, Matera J, Miller MC, Repollet M, Connelly MC, Rao C, Tibbe AGJ, Uhr JW, Terstappen LWMM. Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases. Clin. Cancer Res. 2004;10:6897–904. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15501967. Accessed July 15, 2012.

16. Cohen SJ, Punt CJ a, Iannotti N, Saidman BH, Sabbath KD, Gabrail NY, Picus J, Morse M, Mitchell E, Miller MC, Doyle G V, Tissing H, Terstappen LWMM, Meropol NJ. Relationship of circulating tumor cells to tumor response, progression-free survival, and overall survival in patients with metastatic colorectal cancer. J. Clin. Oncol. 2008;26:3213–21. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18591556. Accessed July 17, 2014.

17. Mikolajczyk SD, Millar LS, Tsinberg P, Coutts SM, Zomorrodi M, Pham T, Bischoff FZ, Pircher TJ. Detection of EpCAM-Negative and Cytokeratin-Negative

Circulating Tumor Cells in Peripheral Blood. J. Oncol. 2011;2011:252361. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3090615&tool=pmcentrez&rendertype=abstract. Accessed August 19, 2014.

18. Nora Dickson M, Tsinberg P, Tang Z, Bischoff FZ, Wilson T, Leonard EF. Efficient capture of circulating tumor cells with a novel immunocytochemical microfluidic device. Biomicrofluidics 2011;5:34119–3411915. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3364832&tool=pmcentrez&rendertype=abstract. Accessed August 19, 2014.

19. Shim S, Stemke-Hale K, Tsimberidou AM, Noshari J, Anderson TE, Gascoyne PRC. Antibody-independent isolation of circulating tumor cells by continuous-flow dielectrophoresis. Biomicrofluidics 2013;7:011807. Available at: http://link.aip.org/link/BIOMGB/v7/i1/p011807/s1&Agg=doi. Accessed December 19, 2013.

20. Xu W, Cao L, Chen L, Li J, Zhang X-F, Qian H-H, Kang X-Y, Zhang Y, Liao J, Shi L-H, Yang Y-F, Wu M-C, Yin Z-F. Isolation of circulating tumor cells in patients with hepatocellular carcinoma using a novel cell separation strategy. Clin. Cancer Res. 2011;17:3783–93. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21527564. Accessed July 25, 2014.

21. Adams DL, Zhu P, Makarova O V., Martin SS, Charpentier M, Chumsri S, Li S, Amstutz P, Tang C-M. The systematic study of circulating tumor cell isolation using lithographic microfilters. RSC Adv. 2014;4:4334. Available at: http://xlink.rsc.org/?DOI=c3ra46839a. Accessed August 19, 2014.

22. Wei L, Jianhong X, Micheli-Tzanakou E. A computational intelligence system for cell classification. Proceedings. 1998 IEEE Int. Conf. Inf. Technol. Appl. Biomed. ITAB '98 (Cat. No.98EX188) 1998:105–109. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=674687.

23. Keenan SJ, Diamond J, McCluggage WG, Bharucha H, Thompson D, Bartels PH, Hamilton PW. An automated machine vision system for the histological grading of cervical intraepithelial neoplasia (CIN). J. Pathol. 2000;192:351–62. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11054719.

24. Wei N, You J, Friehs K, Flaschel E, Nattkemper TW. An In Situ Probe for On-Line Monitoring of Cell Density and Viability on the Basis of Dark Field Microscopy in Conjunction With Image Processing and Supervised Machine Learning. 2007;97:1489–1500.

25. Kachouie NN, Fieguth P, Jervis E. A probabilistic cell model in background corrected image sequences for single cell analysis. Biomed. Eng. Online 2010;9:57. Available at:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2967554&tool=pmcentrez&rendertype=abstract. Accessed September 19, 2012.

26. Hamilton N a, Pantelic RS, Hanson K, Teasdale RD. Fast automated cell phenotype image classification. BMC Bioinformatics 2007;8:110. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1847687&tool=pmcentrez&rendertype=abstract. Accessed June 7, 2013.

27. Phillips KG, Velasco CR, Li J, Kolatkar A, Luttgen M, Bethel K, Duggan B, Kuhn P, McCarty OJT. Optical quantification of cellular mass, volume, and density of circulating tumor cells identified in an ovarian cancer patient. Front. Oncol. 2012;2:72. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3399133&tool=pmcentrez&rendertype=abstract. Accessed March 27, 2013.

28. Phillips KG, Kolatkar A, Rees KJ, Rigg R, Marrinucci D, Luttgen M, Bethel K, Kuhn P, McCarty OJT. Quantification of cellular volume and sub-cellular density fluctuations: comparison of normal peripheral blood cells and circulating tumor cells identified in a breast cancer patient. Front. Oncol. 2012;2:96. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3414893&tool=pmcentrez&rendertype=abstract. Accessed April 4, 2013.

29. Otsu N. A threshold selection method from gray-level histograms. IEEE Trans. Sytems, Man, Cybern. 1979;SMC-9:62–66. Available at: http://web-ext.u-aizu.ac.jp/course/bmclass/documents/otsu1979.pdf. Accessed May 26, 2014.

30. Shamir L, Delaney JD, Orlov N, Eckley DM, Goldberg IG. Pattern recognition software and techniques for biological image analysis. PLoS Comput. Biol. 2010;6:e1000974. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2991255&tool=pmcentrez&rendertype=abstract. Accessed May 27, 2013.

31. Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, Guertin D a, Chang JH, Lindquist R a, Moffat J, Golland P, Sabatini DM. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol. 2006;7:R100. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1794559&tool=pmcentrez&rendertype=abstract. Accessed March 19, 2014.

32. Jones TR, Kang IH, Wheeler DB, Lindquist R a, Papallo A, Sabatini DM, Golland P, Carpenter AE. CellProfiler Analyst: data exploration and analysis software for complex image-based screens. BMC Bioinformatics 2008;9:482. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2614436&tool=pmcentrez&rendertype=abstract. Accessed March 24, 2014.

33. Lamprecht M, Sabatini D, Carpenter A. CellProfiler^{TM}: free, versatile software for automated biological image analysis. Biotechniques 2007;42:71–75. Available at: http://www.biotechniques.com/article/000112257. Accessed May 23, 2014.

34. Shamir L, Orlov N, Eckley DM, Macura T, Johnston J, Goldberg IG. Wndchrm - an open source utility for biological image analysis. Source Code Biol. Med. 2008;3:13. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2478650&tool=pmcentrez&rendertype=abstract. Accessed May 23, 2014.

35. Long F, Peng H, Liu X, Kim SK, Myers E. A 3D digital atlas of C. elegans and its application to single-cell analyses. Nat. Methods 2009;6:667–72. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2882208&tool=pmcentrez&rendertype=abstract. Accessed May 23, 2014.

36. Ozkumur E, Shah AM, Ciciliano JC, Emmink BL, Miyamoto DT, Brachtel E, Yu M, Chen P, Morgan B, Trautwein J, Kimura A, Sengupta S, Stott SL, Karabacak NM, Barber T a, Walsh JR, Smith K, Spuhler PS, Sullivan JP, Lee RJ, Ting DT, Luo X, Shaw AT, Bardia A, Sequist L V, Louis DN, Maheswaran S, Kapur R, Haber D a, Toner M. Inertial focusing for tumor antigen-dependent and -independent sorting of rare circulating tumor cells. Sci. Transl. Med. 2013;5:179ra47. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3760275&tool=pmcentrez&rendertype=abstract. Accessed July 11, 2014.

37. Wang Z, Portier BP, Gruver AM, Bui S, Wang H, Su N, Vo H-T, Ma X-J, Luo Y, Budd GT, Tubbs RR. Automated quantitative RNA in situ hybridization for resolution of equivocal and heterogeneous ERBB2 (HER2) status in invasive breast carcinoma. J. Mol. Diagn. 2013;15:210–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23305906. Accessed August 19, 2014.

38. Scholtens TM, Schreuder F, Ligthart ST, Swennenhuis JF, Greve J, Terstappen LWMM. Automated identification of circulating tumor cells by image cytometry. Cytometry. A 2012;81:138–48. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22170812. Accessed October 22, 2012.

39. Ligthart ST, Coumans F a. W, Bidard F-C, Simkens LHJ, Punt CJ a., de Groot MR, Attard G, de Bono JS, Pierga J-Y, Terstappen LWMM. Circulating Tumor Cells Count and Morphological Features in Breast, Colorectal and Prostate Cancer Katoh M, editor. PLoS One 2013;8:e67148. Available at: http://dx.plos.org/10.1371/journal.pone.0067148. Accessed July 2, 2013.

40. Ligthart ST, Coumans F a W, Attard G, Cassidy AM, de Bono JS, Terstappen LWMM. Unbiased and automated identification of a circulating tumour cell definition that associates with overall survival. PLoS One 2011;6:e27419. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3210171&tool=pmcentrez&rendertype=abstract. Accessed September 11, 2012.

41. Svensson C-M, Krusekopf S, Lücke J, Thilo Figge M. Automated detection of circulating tumor cells with naive bayesian classifiers. Cytometry. A 2014. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24733633. Accessed April 16, 2014.

42. Gleghorn JP, Pratt ED, Denning D, Liu H, Bander NH, Tagawa ST, Nanus DM, Giannakakou P a, Kirby BJ. Capture of circulating tumor cells from whole blood of prostate cancer patients using geometrically enhanced differential immunocapture (GEDI) and a prostate-specific antibody. Lab Chip 2010;10:27–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3031459&tool=pmcentrez&rendertype=abstract. Accessed August 15, 2013.

43. Thege FI, Lannin TB, Saha TN, Tsai S, Kochman ML, Hollingsworth M a, Rhim AD, Kirby BJ. Microfluidic immunocapture of circulating pancreatic cells using parallel EpCAM and MUC1 capture: characterization, optimization and downstream analysis. Lab Chip 2014;14:1775–84. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24681997. Accessed October 4, 2014.

44. Ng A. CS229 Lecture notes Support Vector Machines. :1–25. Available at: http://cs229.stanford.edu/notes/cs229-notes3.pdf.

45. Joachims T. Making large-Scale SVM Learning Practical. In: Scholkopf B, Burges CJC, Smola AJ, editors Adv. Kernel Methods - Support Vector Learn. MIT-Press; 1999. p 41–56. Available at: http://svmlight.joachims.org/. Accessed October 1, 2012.

46. Breiman L. Random forests. Mach. Learn. 2001:5–32. Available at: http://link.springer.com/article/10.1023/A:1010933404324. Accessed May 27, 2014.

47. Jolliffe IT. Principal Component Analysis. 2nd ed. New York, Berlin, Heidelberg: Springer-Verlag New York, Inc.; 2002.

48. Jackson JE. A User's Guide to Principal Components. Hoboken, NJ: John Wiley and Sons; 1991:592.

49. McLachlan G, Peel D. Finite Mixture Models. Hoboken, NJ: John Wiley & Sons, Inc.; 2000.

50. DiCiccio T, Efron B. Bootstrap confidence intervals. Stat. Sci. 1996;11:189–228. Available at: http://www.jstor.org/stable/2246110. Accessed May 27, 2014.

51. Jones TR, Carpenter AE, Lamprecht MR, Moffat J, Silver SJ, Grenier JK, Castoreno AB, Eggert US, Root DE, Golland P, Sabatini DM. Scoring diverse cellular morphologies in image-based screens with iterative feedback and machine learning. Proc. Natl. Acad. Sci. U. S. A. 2009;106:1826–31. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2634799&tool=pmcentrez&rendertype=abstract.

52. Joachims T. Transductive inference for text classification using support vector machines. ICML 1999;0. Available at: http://www1.cs.columbia.edu/~dplewis/candidacy/joachims99transductive.pdf. Accessed May 29, 2014.

2.9 Supplemental Information

2.9.1 Algorithm Details

K-Nearest Neighbors. As introduced by Fix and Hodges, K-nearest neighbors (KNN) is a simple machine learning method in which the class of a new example is determined by a vote of K nearby points in the training set ("nearby" meaning Euclidean distance in the feature space)(1). Cover and Hart validate the theoretical performance of a single-NN algorithm, and show the extent to which "samples which are close together have categories which are close together"(2). Although theoretical bounds exist for K=1, in practice, K is a parameter that may be optimized in training. If K is too small, then the algorithm may be too heavily influenced by a single training example; however, if K is too large, then aggregating many votes may average out real trends. In some cases, particularly if the negative examples greatly outnumber the positive examples in the training data, KNN may perform poorly because the small class nearly always is outvoted. To combat this effect, we introduced a weighting parameter which gave the smaller class an artificial higher voting weight. Although this strategy may give more false positives, it is also likely to reduce the number of false negatives, effectively trading some precision for recall. In our analysis, the number of nearest neighbors, K, and the voting weight were optimized via grid search on leave-one-out validation.

Support Vector Machine. For many binary classification problems, positive and negative examples exist in two distinct regions of the feature space whereby they may be separated by a discriminating hyperplane(3). Support vector machines (SVMs) define this separating hyperplane as a weighted combination of the vectors associated

with the training examples that are closest to the hyperplane. In many classification problems, there is overlap between the classes in the training data, particularly at low signal-to-noise ratio. In these cases, there is no hyperplane that perfectly separates the data. As proposed by Cortes and Vapnik, a regularization parameter may be used, which, when properly optimized, sufficiently smooths over occasional misclassifications and noise in the training data, but not so much that it starts smoothing over the actual information in the training data(4). A support vector machine may be extended into the nonlinear realm through the use of kernel functions(4). For this study, we employ a radial basis function kernel, which, in a transformed space, behaves like a linear hyperplane, but projected back into the original space, effectively grants each support vector a Gaussian-shaped region of influence. The range over which the influence of support vectors decay is governed by a scale parameter. Both the scale and the regularization parameter were optimized together via grid search and cross-validation. For more information, Thorsten Joachims provides a very detailed explanation of support vector machines (5).

Random Forest Classifier. As Breiman outlines (6), a random forest (RF) classifier is a machine learner composed of an ensemble of decision trees. Individual decision trees are a hierarchy of nodes, each of which branches out into more nodes or terminates as a leaf. In each node, there is specified a feature on which the node is splitting and a threshold value (i.e. gate) at which the split occurs. To classify a new example, one follows direction of the splits according to each threshold until a leaf is reached, which specifies the class of the example. This process is analogous to gating in flow cytometry data, in which the population of points is successively split along

threshold values of each feature. When using training data to create such a tree, the algorithm automatically select the proper features on which to split and set the values of the thresholds to obtain the most homogeneous subpopulations(7). This procedure iterates on each sub population until an entire tree is trained with either pure leaves, or smaller than a hand-tuned threshold for the minimum leaf size. CellProfiler makes use of regression stumps (8), which are essentially decision trees with only one split. If one were to, in a random forest, take the limit of the minimum size for a subpopulation as it increases to approach half of the overall number of training examples, one would arrive at such stumps. Thus, by optimizing the minimum size for the split subpopulation, stumps are essentially included the random forest optimization. A second tunable parameter for single trees is an asymmetric weight parameter, which may be used to trade precision for recall.

In comparison to just a single decision tree, many decision trees are trained for a random forest classifier, each on a different bootstrapped sample of the training data. In bootstrapping, a new dataset the same size as the original training set is generated by sampling examples from the training set with replacement(9). Most training examples are included once in the training set, but some are included multiple times and some are skipped. A tree is trained on each bootstrap, so each tree in the ensemble is trained with a slightly different version of the training set. One obtains a single result from the ensemble of trees by aggregating their votes. The bagging process leads to diversity in the trees and robustness against the peculiarities of outliers in the training data [d Richard culter et al ecology]. Often, sufficient diversity in the bagged trees cannot be obtained solely by resampling the data. To elaborate, even on different

bootstrapped samples of the training data, the decision tree training algorithm selects nearly identical features on which to split, and selects nearly identical thresholds for those splits. To combat this phenomenon, at each node for each tree, the random forest algorithm randomly selects only a subset of the features to be available for the decision split [The random subspace method for constructing decision forests tin kam ho]. This strategy ensures that the trees do not become too correlated, each incorporating information from many different features and at different nodes in the tree. Thus, the number of variables available for a split is a third tunable parameter. Because part of the bagging procedure leaves some examples out of the training of individual trees, one can obtain estimates of the outputs of individual trees on out-of-bag examples. These out-of-bag classifications can then be aggregated for validation optimization in place of cross-validation.

Bayesian Classifier with Principal Components Analysis and Gaussian Mixture Modeling. As described in Mitchell's machine learning text (10), a Bayesian classifier chooses the class for a new example ($\hat{y}_{Bayesian\ classifier}$) that maximizes the probability of the class conditioned on the observation ($P(y|\vec{x})$):

$$\hat{y}_{Bayesian\ classifier} = \underset{y}{\operatorname{argmax}}\, P(y|\vec{x}) \tag{1}$$

where $\vec{x}$ is the vector of features and $y \in \{CTC, non\ CTC\}$. Using Bayes' Theorem, this maximization may be reformulated in terms of the priors, $P(y)$, and the class-conditional distributions, $P(\vec{x}|y)$:

$$\hat{y}_{Bayesian\ classifier} = \underset{y}{\operatorname{argmax}}\, P(y)P(\vec{x}|y) \tag{2}$$

For our problem, the priors may be easily estimated by the fractions of CTCs and non-CTCs:

$$P(y = CTC) = \frac{N_{CTC}}{N_{total}};\ P(y = non\ CTC) = \frac{N_{non\ CTC}}{N_{total}} \quad (3;4)$$

The class-conditional distributions, however, require a more sophisticated estimation process which we do via Gaussian mixture modeling (GMM) and principal components analysis to reduce the dimensionality of the feature space. First, the training data is separated by class (i.e. CTC and non-CTC). Then, principal component analysis (PCA) is conducted on each group (11,12). By keeping only the first few columns of the matrix that PCA generates, one may use this modified matrix to project each data vector into a lower-dimensional subspace that is a linear combination of the original features, yet still retains most of the variance of the original data. In the case of image patch pixel intensities as features, because pixels right next to each other retain mostly redundant information, PCA can consolidate most of the information form an entire image patch into a few numbers. In the case of Scholtens' extracted features, a linear combination of features can be more predictive than either one alone, particularly if these features are correlated (e.g. the attributes of the area of a blob and its perimeter). There is no guarantee that such consolidation would occur identically for CTCs and non-CTCs, so the two modified matrices (one derived from CTC examples and one derived from non-CTC examples) are appended to generate a vector that contains the first few principal components in each class. Retaining a higher number of principal components results in a more descriptive subspace, but it also makes over fitting the training data more likely. All of the training data is then projected into this subspace. Then, Gaussian mixture modeling is performed on each group to obtain estimates of the class-conditional distribution functions for CTCs and non-CTCs. GMM will approximate the class-conditional distributions as a weighted

sum of Gaussians (10,13). By using multiple mixands (component Gaussians), multimodal and skewed distributions can be well approximated, as is particularly important for the heterogeneous non-CTC distribution containing debris, leukocytes, and imaging artifacts. The priors of the distribution are estimated by the fractions of total events that are CTC or non-CTC. These priors, however, are reweighted by a tunable parameter to artificially trade precision for recall.

Using the weighted priors and the class-conditional distributions evaluated with the features of the unknown example, the Bayesian classifier selects whichever class has the higher estimated probability. All of the complexity parameters (number of components from CTC PCA, number of components from negative PCA, number of mixands to model CTC class-conditional distribution, number of mixands to model non-CTC class-conditional distribution, prior-inflating parameter, and covariance regularization parameter) were optimized via grid search for F1 score via 3-fold cross validation.

2.9.2 Supplemental Info References

1. Fix E, Hodges JLJ. Discriminatory analysis-nonparametric discrimination: consistency properties. 1951;57:238–247. Available at: http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA800276. Accessed April 28, 2015.

2. Cover T, Hart P. Nearest neighbor pattern classification. Inf. Theory, IEEE Trans. 1967. Available at: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1053964. Accessed April 28, 2015.

3. McLachlan GJ. Discriminant Analysis and Statistical Pattern Recognition. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 1992:1–26. Available at: http://doi.wiley.com/10.1002/0471725293. Accessed April 28, 2015.

4. Cortes C, Vapnik V. Support-vector networks. Mach. Learn. 1995;20:273–297. Available at: http://link.springer.com/10.1007/BF00994018.

5. Joachims T. Making large-Scale SVM Learning Practical. In: Scholkopf B, Burges CJC, Smola AJ, editors Adv. Kernel Methods - Support Vector Learn. MIT-Press; 1999. p 41–56. Available at: http://svmlight.joachims.org/. Accessed October 1, 2012.

6. Breiman L. Random forests. Mach. Learn. 2001:5–32. Available at: http://link.springer.com/article/10.1023/A:1010933404324. Accessed May 27, 2014.

7. Breiman L, Friedman J, Stone CJ, Olshen RA. Classification and Regression Trees. Belmont, California: Wadsworth; 1984.

8. Jones TR, Carpenter AE, Lamprecht MR, Moffat J, Silver SJ, Grenier JK, Castoreno AB, Eggert US, Root DE, Golland P, Sabatini DM. Scoring diverse cellular morphologies in image-based screens with iterative feedback and machine learning. Proc. Natl. Acad. Sci. U. S. A. 2009;106:1826–31. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2634799&tool=pmcentrez&rendertype=abstract.

9. Efron B. Bootstrap methods: another look at the jackknife. Ann. Stat. 1979. Available at: http://www.jstor.org/stable/2958830. Accessed April 28, 2015.

10. Mitchell TM. Machine Learning. WCB McGraw-Hill; 1997.

11. Jolliffe IT. Principal Component Analysis. 2nd ed. New York, Berlin, Heidelberg: Springer-Verlag New York, Inc.; 2002.

12. Jackson JE. A User's Guide to Principal Components. Hoboken, NJ: John Wiley and Sons; 1991:592.

13. McLachlan G, Peel D. Finite Mixture Models. Hoboken, NJ: John Wiley & Sons, Inc.; 2000.

CHAPTER 3

Automated electrorotation shows that electrokinetic separation frequency window in pancreatic cancer cells is robust to acquired resistance to chemotherapy, serum starvation, and epithelial to mesenchymal transition. [2]

3.1 Abstract

The capture of circulating tumor cells via immune-affinity may be compromised by reduced antigen expression associated with acquired resistance to chemotherapy, deprivation growth factors when entering circulation, and the epithelial-to-mesenchymal transition. Dielectrophoresis (DEP) has been shown to enhance capture of model cell lines in microfluidic devices, but before DEP-enhanced capture may be used for cells in circulation, robustness of cell electrical properties must be known after being perturbed to stimuli that would be present in circulation. To this end, we used automated electrorotation to measure the cytoplasmic permittivity, cytoplasmic conductivity, and specific membrane capacitance of pancreatic cancer cells under three treatments. First, we developed two gemcitabine-resistant sub-clones of BxPC3 pancreatic cancer cells and compared them to gemcitabine-naive parental cells. Second, we serum starved BxPC3 and PANC-1 cells and compared them to untreated counterparts. Third, we induced the epithelial-to-mesenchymal transition in PANC-1 cells and compared them to untreated PANC-1 cells. We also measured electrorotation

[2] Parts of this chapter have been submitted for publication: Changes to electrical properties of pancreatic cancer cells due to acquired chemotherapy resistance and epithelial-to-mesenchymal transition revealed via Electrorotation. Timothy Lannin, Wey-Wey Su, Conor Gruber, Ian Cardle, Chao Huang, Fredrik Thege, and Brian Kirby. *Biomicrofluidics*.

spectra of white blood cells isolated from a healthy donor. The properties from fit electrorotation spectra were used to compute DEP spectra and crossover frequencies. For all three experiments, the median crossover frequency for both treated and untreated pancreatic cancer cells remained significantly lower than the median crossover frequency for white blood cells. The robustness of the crossover frequency to these treatments indicates that DEP is a promising technique for enhancing capture of circulating cancer cells.

3.2 Introduction

Cancer is a leading cause of death, a statistic that is made worse by many cases of late detection, often after the onset of metastasis and its characteristic spread of cancerous cells away from the primary tumor to form new tumors throughout the body. An expected mediator in this process is circulating tumor cells (CTCs), which are cancer cells found in patient blood at an extremely low concentration, about 1-10 CTCs per ml of blood (1). These cells originate from tumors and a subpopulation has been shown to be pre-cursors to metastasis (2). In addition to this putative role, it is well known that these circulating tumor cells correlate with disease progression (3), that they can be captured even in early disease (4), and that they can be used to gauge functional cell responses to pharmaceuticals (5) or the genetic character of the cancer (6). This information from captured CTCs is valuable to both clinicians and researchers particularly in the context of pancreatic cancer, in which the presence of circulating pancreatic cells precedes tumor formation(2). Furthermore, circulating

pancreatic cells exist in a subset of patients without cancer but with potentially cancerous pancreatic cystic lesions(4).

To obtain this information, CTCs have been successfully captured from blood samples through mechanical (7), electromechanical (8) and immunoaffinity methods (9–11), the last being the most common. However, challenges exist capturing subsets of CTCs that have reduced antigen expression, often associated with the epithelial to mesenchymal transition (EMT)(12) or acquired resistance to chemotherapy(13). Also, performing more complex and informative procedures requires a higher purity of captured cells, and improvements to the current state-of-the-art are needed.

This improvement could be achieved through dielectrophoresis (DEP). The use of nonuniform alternating current electric fields can be used to attract and repel objects with different complex permittivity values depending on the field frequency. Using DEP as the only separation mechanism has led to some success in CTC enrichment(14–16). A proof-of-concept hybrid DEP-immunoaffinity devices has shown promise of further improving capture(17–19). Careful device design and frequency selection can attract CTCs while repelling contaminating white blood cells, and simulations have predicted capture efficiency increases of up to 400% (20).

The upper half of Figure 3.1 shows measured DEP spectra for different white blood cells and cancer cell lines. Differences in the crossover frequency, the frequency in which a cell changes from exhibiting negative to positive DEP, between cancer cells and white blood cells permit positive selection of the rare cancer cells within a blood sample. This can be done by operating DEP at an optimal frequency between the two different crossover frequencies.

Unfortunately, measuring DEP forces against frequency in order to measure accurate DEP spectra can be experimentally difficult, because in order to estimate DEP forces, one must carefully model the electric field intensity and the hydrodynamics of cells as the two effects vary spatially in the device. Furthermore, the magnitude of negative DEP is difficult to measure because cells are forced to regions in the device where the field is weak, so DEP trapping experiments yield only partial spectra(21). Additionally, near the crossover frequency, which is of greatest interest for cell separations, the DEP signal is necessarily weakest, making it more vulnerable to confounding effects.

A related technique, electrorotation (ROT), relates to DEP forces through the Clausius-Mossotti factor(22–24). In ROT, a nearly uniform rotating electric field induces constant torques on (and thus steady rotation rates of) the cells. By measuring cell rotation rate as a function of applied field frequency, one may trace out a ROT spectrum. This spectrum is proportional to the imaginary part of the cell's Clausius-Mossotti factor, which contains all of the information necessary to reconstruct the real part through either the Kramers–Krönig relations or by fitting cells' dielectric properties to a multi-shell model (22–24). Corresponding DEP-ROT plots are shown in Figure 3.1.

Using cell lines exclusively cannot be a direct substitute for capture models for CTCs in the bloodstream because of possible changes in phenotype in these cells. Changes in CTC size, surface markers, and nuclear-to-cytoplasm ratio can occur due to vesicle shedding processes (25), the epithelial-to-mesenchymal transition (EMT) (9,11,26) in which tumor cells can be reprogrammed as cancer develops, or developed

cancer therapy resistance (27). Cell lines have been shown to be too large and express too much surface marker (28,29) relative to bloodstream-captured CTCs. Such acquired differences can also result in different DEP spectra, inducing shifts in the crossover frequency.

An investigation into these different variants of developmental differences is therefore necessary in order to gain an understanding of the whole picture behind the characterization of CTC electrical properties. In this study, we measured electrorotation spectra of gemcitabine-resistant, serum starved, and EMT-induced pancreatic cancer cell lines in addition to the parent/untreated cells and healthy donor white blood cells in order to fit cell electrical properties and compute the full effects on DEP spectra. These measurements can then inform optimal settings for DEP-enhanced cell capture devices.

3.3 Materials and Methods

3.3.1 Cell Culture and Preparation

A low-conductivity isotonic sugar buffer used as the medium for electrorotation was prepared with 9.5% sucrose, 0.3% dextrose, 0.1% Pluronic F68, 0.1% BSA purchased from Sigma-Aldrich (St. Louis, MO) in de-ionized water, with PBS added until the conductivity was 70 mS/m (with automated temperature compensation to 20C), approximately 2 mL of PBS per 50 mL of buffer. This ROT buffer was refrigerated at 4C and warmed to 37C before experimentation. Modeled as a pure sucrose solution, the media was assumed to have a relative permittivity of 74.2 (30).

For experimentation, 70% confluent adhesion cultures and suspension cultures kept at 3×10^5 to 2×10^6 cells per mL were cultured in Corning CellBIND surface 25cm^2 flasks from ThermoFisher Scientific (Waltham, MA) with 5 mL of cell culture media supplemented 10% with fetal bovine serum and 100 units penicillin, 0.10 mg streptomycin and 0.25 g amphotericin B per ml in a 5% CO_2, 37C humidified incubator. Cell cultures included BxPC3 pancreatic, PANC-1 pancreatic, and suspension U937 lymphoma human cancer cells from the American Type Culture Collection (ATCC; Manassass, VA)?. BxPC3 and U937 cells were cultured with Roswell Park Memorial Institute-1640 (RPMI) purchased from Lonza (Basel, Switzerland), while PANC-1 cells were cultured from Dulbecco's Modified Eagle Medium (DMEM) media purchased from Mediatech (Manassass, VA).

Gemcitabine-resistant subclones were isolated by exposing gemcitabine-sensitive parental BxPC-3 cells to increasing concentrations of gemcitabine over a 10-month period. Clones were isolated at 80nM and 360nM gemcitabine, referred to as BxGR-80C and BxGR-360C respectively. Resistance was confirmed by measuring and comparing gemcitabine IC_{50} values. These sub-clones are discussed in detail elsewhere (cite manuscript in preparation/submission).

For the serum starvation studies, adhesion cell cultures of PANC-1 or BxPC3 cell lines were seeded into two CellBIND flasks 48 hours before performing electrorotation studies. One flask would be passaged with regular media with both the 10% serum and 1% antibiotics, while the other flask was passaged using media with only the 1% antibiotics. We would then perform an electrorotation experiment on both samples, one from each flask.

EMT induction was performed by plating PANC-1 cells and controls at about 6.25e3 cells per square centimeter in 6-well plates approximately six days before the scheduled experiment. We then treated the cultures with 10 ng/ml TGF-beta and 20ng/ml EGF in reduced serum conditions (2.5% FBS RPMI-1640) on days 1, 3, and 5 before performing electrorotation experiments on the sixth day on both the EMT induced cells and the control cells.

Adhesion cultures were rinsed with PBS twice, then lifted off by treating in trypsin for 5 minutes. The suspension was washed three times by centrifuging at 300xg for 5 minutes, removing the supernatant with a pipette, and re-suspending the cell pellet in 1mL ROT buffer. Optionally, the cell suspension was then diluted with additional ROT buffer to reduce the number of cells in the field of view to approximately 15.

3.3.2 Experimental Apparatus

An electrorotation cell with hyperbolic 4-electrode geometry (31) was fabricated with Chromium-Gold-Chromium electrodes on 0.5mm-thick borofloat glass wafers in the Cornell Nanoscale Science and Technology Facility (Ithaca, NY) using standard photolithographic techniques (described in detail in (18)). The tip-to-tip distance between opposite electrodes was 800 μm (similar to (32)), and the hyperbolae extended until the narrowing gap between adjacent electrodes was 133 μm (Figure 3.2). This gap was sufficiently large such that the device had large impedance compared to the 50Ω terminating resistors in parallel with each electrode connection. A 5mm-diameter plastic well was epoxied around the electrode gap to hold the cell suspension. Four sinusoidal voltages in quadrature (all 3.8 $V_{p\text{-}p}$) were applied to equal

length 50Ω coaxial cables via a four-channel function generator (ArbStudio 1104, Teledyne Lecroy).

3.3.3 Capturing ROT Videos

To conduct an electrorotation run, cell suspension (30μL) was pipetted onto into the well, and cells settled for approximately 30 seconds. During the settling period, the function generator applied 51.4 kHz ROT signal continuously to prevent the settled cells from adhering to the glass and to reach a temperature steady-state before the start of the ROT data acquisition. We positioned the microscope stage so the field of view was in the approximate center of the electrodes where the field is uniform(33). For an electrorotation run, we programmed five-second intervals of 36 different frequencies equally spaced on the log scale over the range 1.125kHz-62.5kHz into the ArbStudio software, with randomized order to decorrelate the applied frequency with any time-varying effects on the experiment (e.g. drift of cells into varying-strength field, electrical properties varying with temperature, etc.) (Figure 3.2). This sequence was preceded by a brief pause, then 5 seconds of CW field, and then 5 seconds of CCW field (both at 51.4kHz) in order to synchronize the video with a clearly-defined pattern in the applied voltage sequence.

We illuminated the cells under brightfield with the condenser aperture wide open in order to minimize exposure times to maximize frame rate. An Eclipse TE2000U inverted microscope by Nikon (Melville, NY) equipped with a 40x/0.6 Plan Fluor objective, a RETIGA EXi FAST (8-bit images, 1392x1040 pixels, no binning), and NIS Elements D software captured a time-lapse of images. We prescribed frame rates

at 10 FPS, but camera readout speed limited actual acquisition to ~7 FPS. After video was captured, we rinsed the well on the ROT device at least three times with deionized water before proceeding with a new sample.

3.3.4 Acquisition of Individual Cell Rotation Rates From Videos

NIS Elements software exported video frames were exported to .tif files, which MATLAB imported for further processing. Manually estimating the rotation rate of the cells is quite tedious, so we developed an algorithm, similar to some found in the literature (34–36), to automate parts of the rotation rate estimation.

First, for each pixel spatial location in the video, the script subtracted the background by computing the temporal median brightness and subtracting it from each video frame to suppress debris in the microscope optics and in the ROT chamber and nonuniformity in the illumination that could later interfere with image-based automated rotation estimation.

Second, because cells drifted throughout the course of an experiment, the script needed a frame of reference to translate an image patch, centered on the cell of interest, in order decouple translation and rotation. A user manually entered an initial guess for the trajectory of the cell in a custom MATLAB GUI by clicking on the cell center periodically throughout the course of the video and interpolating to acquire intermediate approximate cell center positions. The script used these positions to seed an optimization that attempted to maximize the correlation between two image patches by finding the translation and rotation of the cell between the two patches.

Third, the script ran optimization routine to compute the frame-to-frame rotation rate (as well as small translations of the cell about its initial guess). The objective function is defined as follows. Given an (x,y) position (not necessarily integer pixels) of the center of the cell at one time (frame 1), interpolate between pixels to extract an image patch, P1, (27px*27px) centered on that cell. Also, given a guess about the translation and rotation of the cell between this initial frame and the frame at a later time (frame 2), compute a set of translated-and-rotated pixel positions, and interpolate these positions on frame 2 to generate a second image patch, P2. To both patches, apply a blurred-edge circular mask (radius 7px, Gaussian blur radius 3px) to suppress the corners of the patches and include only pixels within the cell in subsequent steps. Compute the digital image correlation between the two masked patches (mP1 and mP2):

$$\rho = \frac{\sum_i \sum_j [mP1(i,j) - \overline{P1}][mP2(i,j) - \overline{mP2}]}{\sqrt{(\sum_i \sum_j [mP1(i,j) - \overline{mP1}]^2)(\sum_i \sum_j [P2(i,j) - \overline{mP2}]^2)}} \quad (1)$$

where (i,j) is a pixel within the patch, and $\overline{mP1}$ and $\overline{mP2}$ are the mean intensities of P1 and P2. Thus, the objective function optimizes over the arguments position and rotation, (Δx, Δy, θ), which determine the set of locations to interpolate from frame 2. MATLAB's fmincon tool minimized the negative of this objective function. We set tolerance in the change of magnitude of the vector (Δx, Δy, θ) [px, px, degrees] to 10^{-6}. The initial guess for x and y originated from the manual clicking, and the initial guess for θ originated from the optimization of the previous two frames. Optimization constraints for translations were ±5px about the guess value supplied by manual clicking and for rotation was ±49° about the initial guess. Because fmincon is a

gradient-based optimizer, it needed cubic interpolation of the image intensity in order to make the objective function sufficiently smooth for proper convergence.

We have used a small applied voltage and a large electrode gap, leading to smaller applied field magnitudes and cell rotation speeds. Although this low field strength has the advantage of suppressing joule heating, cell rotation rates were so slow that cell rotation between adjacent frames was often sub-pixel, so angle displacements were compared over time increments of ~1 second, rather than 1 frame.

The final step of the rotation rate estimation is to break apart the ROT rate versus time over the whole video, as estimated from the angle displacement optimization at each frame, into the typical ROT rate within the 5-second window that each frequency was applied. We accomplished this synchronization by manually selecting a characteristic sharp peak at the beginning of a plot of accumulated angle versus time. We then took the median of all of the (potentially noisy) ROT rate measurements for each pair of frames in this each 5-second window as an outlier-resistant measure of the average ROT rate at that frequency. This automated ROT estimation correlated well with manual ROT rate estimates from a custom MATLAB GUI in which a user clicked on a distinctive feature on a cell, tracking the feature as the cell rotated throughout the video (Figure 3.3).

3.3.5 Analytical Model of ROT Spectra

Fundamental equations for dielectrophoresis and electrorotation are described thoroughly in the literature (22–24,37). Briefly, complex permittivity depends on the

substance permittivity, ε, conductivity, σ, and the angular frequency of the applied field, ω:

$$\underline{\varepsilon} = \varepsilon + \frac{\sigma}{j\,\omega} \tag{2}$$

A single spherical uncharged particle with homogeneous, isotropic conductivity and permittivity in a uniform sinusoidal external electric perturbs the field equivalent to a dipole with moment phasor:

$$\underline{\vec{\boldsymbol{p}}_{\mathbf{0}}} = 4\pi\,\varepsilon_m \vec{\boldsymbol{E}}_{\mathbf{0}}\,a^3\,\frac{\underline{\varepsilon_p} - \underline{\varepsilon_m}}{\underline{\varepsilon_p} + 2\underline{\varepsilon_m}} = 4\pi\,\varepsilon_m \vec{\boldsymbol{E}}_{\mathbf{0}}\,a^3\,\underline{f_{CM}} \tag{3}$$

where ε_m is the medium permittivity, $\vec{\boldsymbol{E}}_{\mathbf{0}}$ is the external field phasor, a is the particle radius, $\underline{\varepsilon_p}$ is the particle effective complex permittivity, $\underline{\varepsilon_m}$ is the medium complex permittivity, and $\underline{f_{CM}}$ is the complex Clausius-Mossotti factor. If the sphere is not homogenous, but rather made of a thin membrane of zero conductivity and low permittivity compared to its contents (e.g. cytoplasm), the effective complex permittivity of that particle is

$$\underline{\varepsilon_p} = \frac{\underline{\varepsilon_{cyto}}\,C''a}{\underline{\varepsilon_{cyto}} + C''a} \tag{4}$$

where C'' is the membrane capacitance per unit area and $\underline{\varepsilon_{cyto}}$ is the complex permittivity of the contents. This dipole interacts with a nonuniform external field to give rise to the familiar time-averaged dielectrophoresis force:

$$\langle \vec{\boldsymbol{F}}_{DEP} \rangle = \pi\,\varepsilon_m\,a^3\,\mathrm{Re}\left(\underline{f_{CM}}\right)\,\nabla(\vec{\boldsymbol{E}}_{\mathbf{0}} \cdot \vec{\boldsymbol{E}}_{\mathbf{0}}) \tag{5}$$

which is proportional to the real part of the Clausius-Mossotti factor. The dipole may also interact with a rotating external field to yield a time-averaged torque:

$$\langle \vec{\boldsymbol{T}}_{ROT} \rangle = -4\pi\, \varepsilon_m\, a^3\, \mathrm{Im}\left(\underline{f_{CM}}\right) \left[\mathrm{Re}\left(\underline{\vec{\boldsymbol{E}}_{\boldsymbol{0}}}\right) \times \mathrm{Im}\left(\underline{\vec{\boldsymbol{E}}_{\boldsymbol{0}}}\right)\right] \tag{6}$$

which is proportional to the negative of the imaginary part of the Clausius-Mossotti factor. In the Stokes flow limit, the viscous torque, proportional to cell rotation rate, instantly balances the ROT torque. Thus, the imaginary part of the Clausius-Mossotti factor (which varies with applied field frequency) is proportional to the observed cell rotation rate at that applied frequency:

$$\mathrm{Im}\left(\underline{f_{CM}(\omega)}\right) \propto \Omega(\omega) \tag{7}$$

With a detailed model of the applied field strength and the viscous hydrodynamics, we could directly solve for $\mathrm{Im}\left(\underline{f_{CM}}\right)$, but such modeling challenges may be avoided by including a scale factor accounting for field strength and the viscous hydrodynamics in the curve fit.

3.3.6 Robust Parameter Fits to ROT Measured ROT Spectra

In the inertia-free limit of Stokes flow, ROT rates are proportional to the imaginary part of the Clausius-Mossotti factor. Rather than attempting to model the complicated friction between the cell and the wall and the precise magnitude of the electric field at each point in the ROT chamber, we fit ROT spectra to $\mathrm{Im}(f_{CM})$ up to a scale factor, which accounted for field strength and friction. In addition to the scale factor, the parameters also used make the fit are the specific membrane capacitance, the cytoplasmic conductivity, and the cytoplasmic permittivity.

Despite carefully designing an image processing algorithm and using the median within each frequency window to suppress outliers, the measured ROT spectra occasionally contain a few outliers. These outliers can strongly influence curve fits

(Figure 3.4), so nonlinear least trimmed squares was selected as an outlier-resistant curve fitting scheme (38). To conduct least trimmed squares, an initial guess of parameters generates a model ROT rate for each frequency, $\Omega_{model}(\omega_i, params)$. The squared deviation of the model from the data is then computed for each point in the spectrum:

$$g_i = (\Omega_{meas}(\omega_i) - \Omega_{model}(\omega_i, params))^2 \quad (8)$$

a priori, the user had decided some number of the points to trim from the fit (3 for our work). Points that deviate the most are then trimmed from the dataset, and a new model is computed via standard nonlinear least squares. A new set of g_i's are computed (now including the trimmed points). This process iterates until the same set of points are trimmed between iterations.

We used MATLAB's fmincon to do the intermediate nonlinear least squares fit, but rather than fitting σ_{cyto}, ε_{cyto}, C''_{mem}, and, directly, the logs of these quantities were fit to the curve. This ensured that 1) the physical quantities always remained positive, 2) the quantities were all of a similar order of magnitude, and 3) the sensitivity of fit the parameters to the data varies on a log scale as the data points are spaced in log frequency. The scale parameter was not fit on a log scale. The guess used to initialize all parameters were: scale = -0.5269 rad/s, σ_{cyto} = 0.5299 S/m, ε_{cyto} = 93.1, and C''_{mem} = 19.9 mF/m^2, which were median parameters computed from an initial optimization of the data. Constraints on the parameters were scale ∈ [-10,10] rad/s, σ_{cyto} ∈ [0.1, 100] S/m, ε_{cyto} ∈ [10, 1000], and C''_{mem} ∈ [0.15, 1500] mF/m^2, well beyond the physical range of these parameters that one would expect from the literature. We excluded any

cell whose optimization converged such that a property was at a constraint an outlier, and omitted from further analysis.

3.4 Results and Discussion

Our device applied rotating electric fields of various frequencies to cancer cells and white blood cells, and an image processing script estimated subsequent rotation rates of cells automatically from the captured video. Because of the cells' minimal inertia, viscous forces dominate and their rotation rate is proportional to the electrorotation torque. The electrorotation torque is proportional to the imaginary part of the Clausius-Mossotti factor, which depends on the permittivity and conductivities of the cell cytoplasm and the specific capacitance of the cell membrane. With an additional measurement of the radius of the cell measured from the bright field images, we inferred these cell electrical properties (along with a proportionality factor accounting for electric field magnitude and viscous torques) from curve fits of the cells' measured ROT spectra. ROT measurements are, by nature, single cell, enabling not only measurements of average cell properties, but also quantification of the distribution of cell properties. Figure 3.5 highlights the spread of fit ROT spectra, allowing for a comparison of cells within each population. Most cells consistent with two relaxations that appear as a peak and a trough in the ROT spectra: 2) a low frequency (~10^5 Hz) peak dominated by the cell radius, specific membrane capacitance, and suspending medium conductivity, and 2) high-frequency (~10^6-10^8 Hz) trough associated with DEP transitioning from being dominated by the cell-medium conductivity mismatch to the cell-medium permittivity mismatch.

Figure 3.6 shows the effect of acquired resistance to gemcitabine on BxPC3 pancreatic cancer cell sub-clones. Most notably for the design of future DEP devices, the calculated crossover frequency significantly decreased for both gemcitabine-resistant sub-clones as compared to gemcitabine-naïve parental BxPC3s ($p<0.05$, Wilcoxon rank sum test), effectively widening the frequency window at which a majority of cancer cells experience pDEP and a majority of WBCs experience nDEP. Because the time constant associated with this relaxation is approximately proportional to the product of specific membrane capacitance and cell radius, the observed decrease in crossover frequency is explainable due to the observed increase in median cell radius and in median specific membrane capacitance of the gemcitabine-resistant sub-clones (Table 1). If the trend of a larger specific membrane capacitance for gemcitabine-resistant cancer cells also exists *in vivo*, then separation potential would exist even noting that (for prostate cancer) CTCs tend to be smaller than their cell line counterparts (39). These gemcitabine-resistant clones developed one of many possible resistance mechanisms. It is possible that clones that have acquired resistance via different mechanisms may exhibit gross physical changes that have a less favorable impact on their crossover frequency in terms of remaining separable from blood cells. Studying such clones will be the subject of future work.

Figure 3.7 shows the effect of serum starvation on BxPC3 and PANC1 cells. Serum-starving PANC1 cells significantly increased their crossover frequency, though the crossover frequency still remains significantly higher than that of WBCs (see Table 1, $p<0.05$, Wilcoxon rank sum tests). No significant change in crossover frequency was observed for serum-starved BxPC3 cells as compared to their non-

starved controls, despite the duration of serum starvation (48 hours) being longer than protocols used to cause vesicle blebbing from the cell membrane (overnight)(40) or longer than the maximum lifetime of cancer cells in circulation (<24 hours)(41). The lack of an observed change in crossover frequency is likely due to our observation that cell radii tended to increase with serum starvation while specific membrane capacitance tended to decrease.

We implemented a well-used EMT protocol, similar to work from the literature(42). Our EMT-induced cells had spindle-like morphology and a pre-liftoff surface area that was larger the untreated PANC1 cell surface area (Figure 3.8, h-i), as well as EpCAM downregulation, cytokeratin downregulation, E-cadherin downregulation, N-cadherin upregulation, and vimentin upregulation (as confirmed by immunofluorescence, data not shown). Furthermore, Wang et al. (13) observed a strong connection between gemcitabine resistance in pancreatic cancer cells and the acquisition of mesenchymal characteristics (e.g. an increase in pre-liftoff surface area). Thus, we had expected find a strong difference between EMT-induced PANC1 cells and their untreated controls, because as a consequence of this morphology change. According to Wang et al.(43), Shim et al. (8), and Gascoyne et al. (44), this increase in surface area would putatively change the ruffles and folds in the cell membrane, increasing their effective surface area and specific membrane capacitance. Furthermore, Salmanzadeh et al. (45) and Mulhall et al. (46) observed that aggressiveness of cancer cell phenotype correlates with an increase in specific membrane capacitance for mouse ovarian surface epithelial and oral cancer cells, respectively. We, however, failed to observe such a change on the electrical properties

of EMT-induced PANC1 cells. Figure 3.8 (a-g) show the effect of EMT induction on the DEP spectra and electrical properties of PANC1 cells. No significant change was observed in the crossover frequency of the EMT-induced cells compared to their non-induced counterparts (Table 1). However, regardless of EMT treatment, PANC1 cells' crossover frequency remained significantly higher than that of WBCs.

ROT measurements enable simultaneous fits of multiple cell properties, compared to crossover frequency measurements allow for fitting of only one parameter (typically specific membrane capacitance). Because of this increased information, we were able to make estimates of the cells' cytoplasmic relative permittivity. White blood cell median permittivity was measured to be 111 (95% bootstrapped confidence interval 79–122), and typical median pancreatic cancer cell relative permittivity was measured to be around 96 (see Table 1). These values are higher than those from ROT experiments of T lymphocytes and MDA321 breast cancer cells (64 and 52) reported by Becker et al. (16) and used by many subsequent studies (14,17,18,47). Because the fit relative permittivity is higher than that of the suspending medium (74.2), these data predict that no upper crossover frequency exists for these cells. Unfortunately, the upper crossover frequency predicted by the low permittivity values in the literature was higher than our experimental apparatus can generate, so we could not validate this prediction.

3.5 Conclusion

By using rotating electric fields to induce frequency-dependent-torques on polarizable particles (electrorotation), we have inferred the electrical properties

(cytoplasm permittivity and conductivity as well as cell membrane specific capacitance) of pancreatic cancer cells and blood cells by their observed rotation rates at each field frequency. Electrorotation measurements are by nature single-cell, and this study has revealed the distribution of electrical properties within cancer cell line populations. For all pancreatic cancer cells measured and under all measured treatments of the cells (acquired resistance to gemcitabine, serum starvation, and induced epithelial-to-mesenchymal transition), a significant window exists within which at least a majority of cancer cells exhibit positive DEP while a majority of blood cells exhibit negative DEP. The robustness of the distinguishing electrical properties to the applied treatments is promising for DEP-enhanced cell capture devices, as proposed in other work (17,18,20). Furthermore, these measurements revealed a higher cytoplasmic permittivity than has been previously reported in the literature, which predicts that no upper crossover frequency exists and could be the study of future investigation. Future work could involve studying ROT of cancer cells at higher frequencies to validate this measurement and DEP at higher frequencies to exploit permittivity mismatches to separate cells. Additional future work could involve trying single cell expression levels of various markers (e.g. for EMT) and the corresponding cells' ROT spectra by taking an immunofluorescence image prior to capturing ROT videos, which would strengthen the connection between single cell electrical properties and expression levels of markers of interest

3.6 Acknowledgements

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. (DGE-1144153) (TBL). The authors would like to thanks Cornell University's Engineering Learning Initiatives (WWS), HHMI med-into-grad scholarship (FIT), the Lester and Sheila Robbins Scandinavian Graduate Student Fellowship (FIT), for financial support. This work was performed in part at the Cornell NanoScale Facility, a member of the National Nanotechnology Coordinated Infrastructure (NNCI), which is supported by the National Science Foundation (Grant ECCS-1542081). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

3.7 Figures, Tables, and their Legends

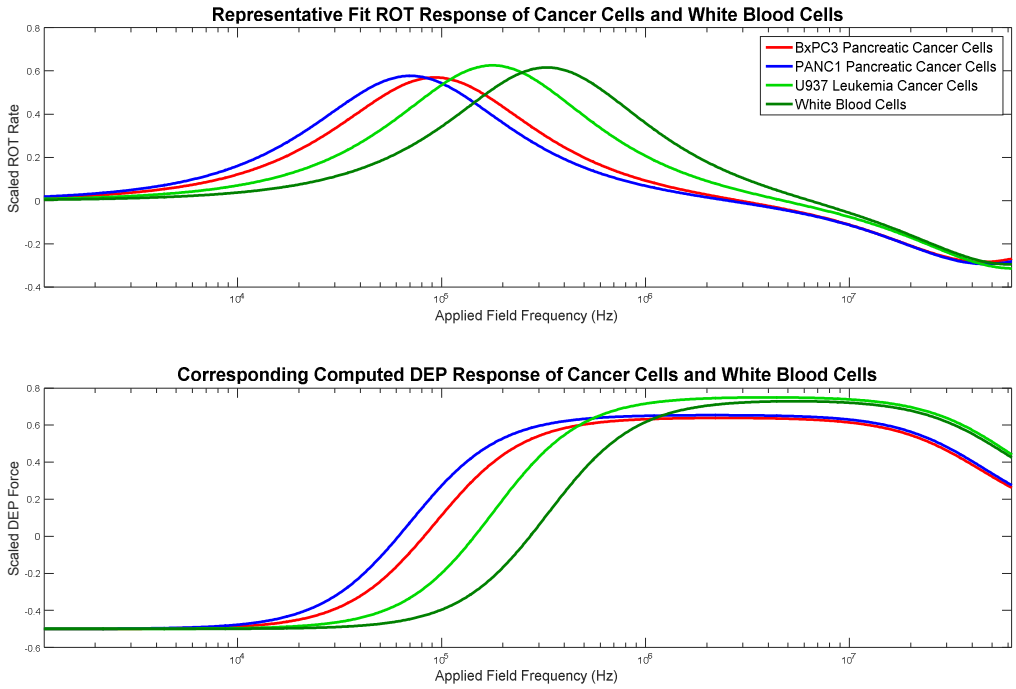

Figure 3.1 – Calculated DEP and ROT Spectra of untreated PANC-1 and BxPC3 pancreatic cancer cells compared to that of U937 leukemia blood cells and white-blood cells from healthy donors. Curves were generated by taking the median of the fit cell electrical properties of the corresponding populations. Differences in the crossover frequency between the cancer cells and the white blood cells allow for positive DEP selection.

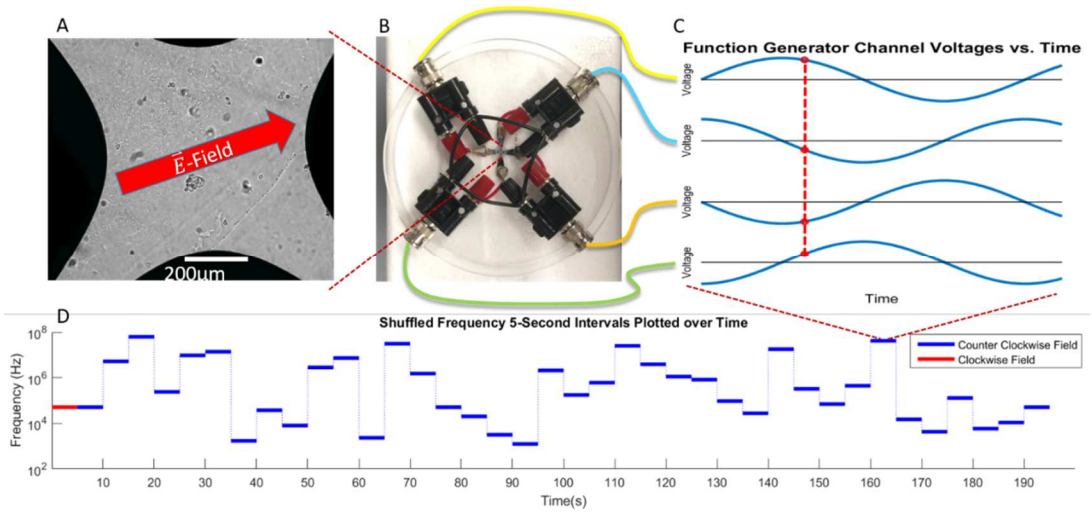

Figure 3.2 – Schematic of the overall ROT device and operation, with plotted sinusoidal voltages and frequencies as shown. a) A 10x magnified view of the electrodes in the ROT device, with an electric field direction shown corresponding to the voltages at the marked time frame shown in C. b) Overhead view of the ROT device, with electrodes silver epoxied to red wires in order to connect to mutually grounded BNC breakout connectors. c) Sinusoidal voltage functions within each five second interval, supplied by the function generator to the separate connectors, each with a different 90-degree delay from the last one in order to generate an electric field in the direction shown. This configuration would result in a counterclockwise rotating electric field vector. d) Shuffled frequency five-second intervals over the entire course of a single ROT experiment. These were randomized in order to eliminate any possible dependence of experimental time on rotation rate (e.g. Joule heating or drifting of cell position).

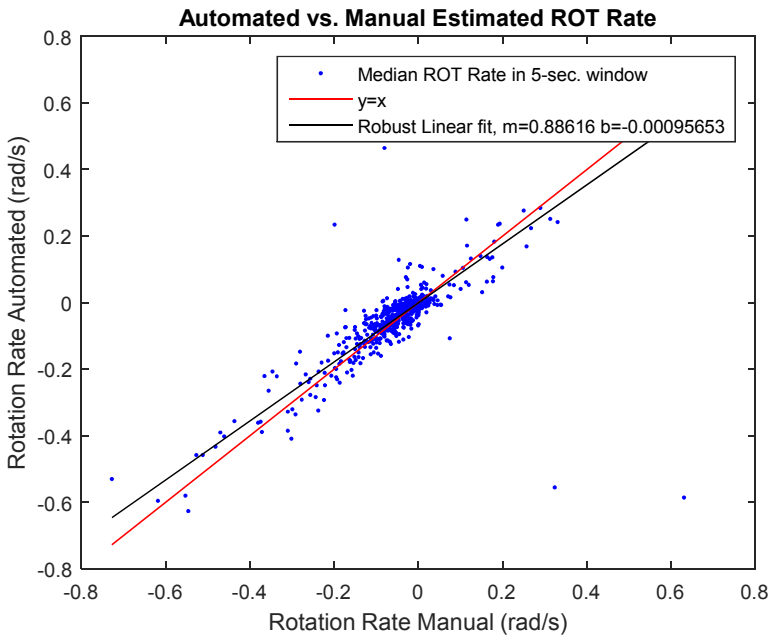

Figure 3.3 – Estimated rotation rate of cells as measured by the automated image processing algorithm versus the rotation rate as measured by manually tracking features. Within each 5-second window that a particular field frequency was applied, approximately 37 frames were captured yielding multiple measurements of the ROT rate. Each dot represents the median of such ROT rates in the window. The automated ROT rate estimation correlates strongly with the manual measurements, though slightly under predicting the manual ROT rate (least trimmed squares fit slope = 0.88 vs. ideal slope = 1).

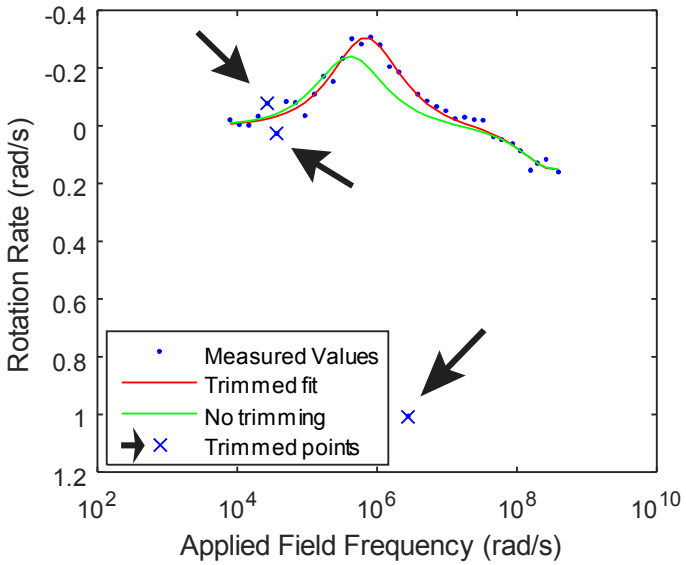

Figure 3.4 – A least trimmed squares fit and a traditional nonlinear least squares fit showing the influence of outliers on the two curve fits on a ROT spectrum. Note the direction of the y-axis is reversed to make this plot of rotation rate consistent with the imaginary part of the Classius-Mossotti factor, which as the opposite sign.

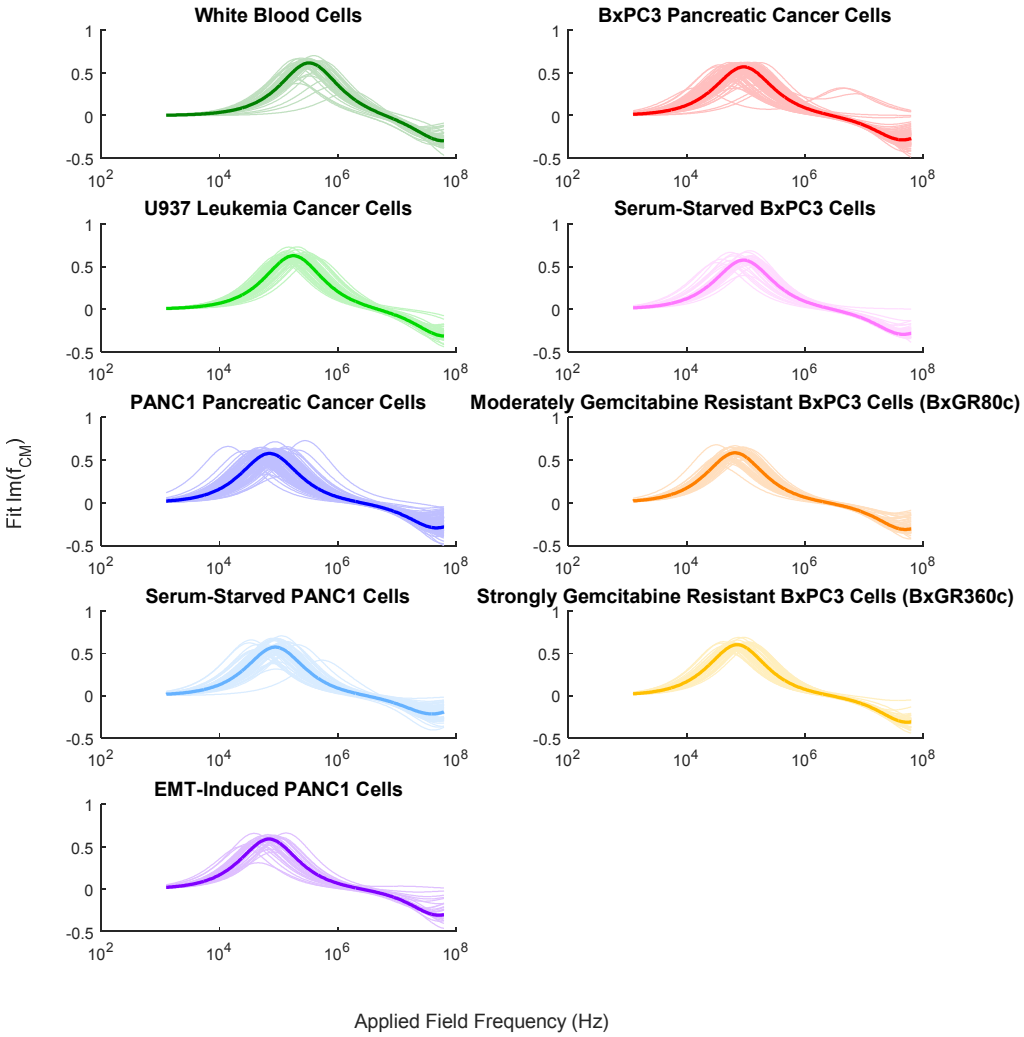

Figure 3.5 – Fit electrorotation spectra. Each faint curve depicts the imaginary part of the Classius-Mossotti factor fit to an individual cell of a particular type. Bold curves represent a theoretical cell with cytoplasmic conductivity, cytoplasmic permittivity, membrane specific capacitance, and radius equal to the median fit values for the cells of that type.

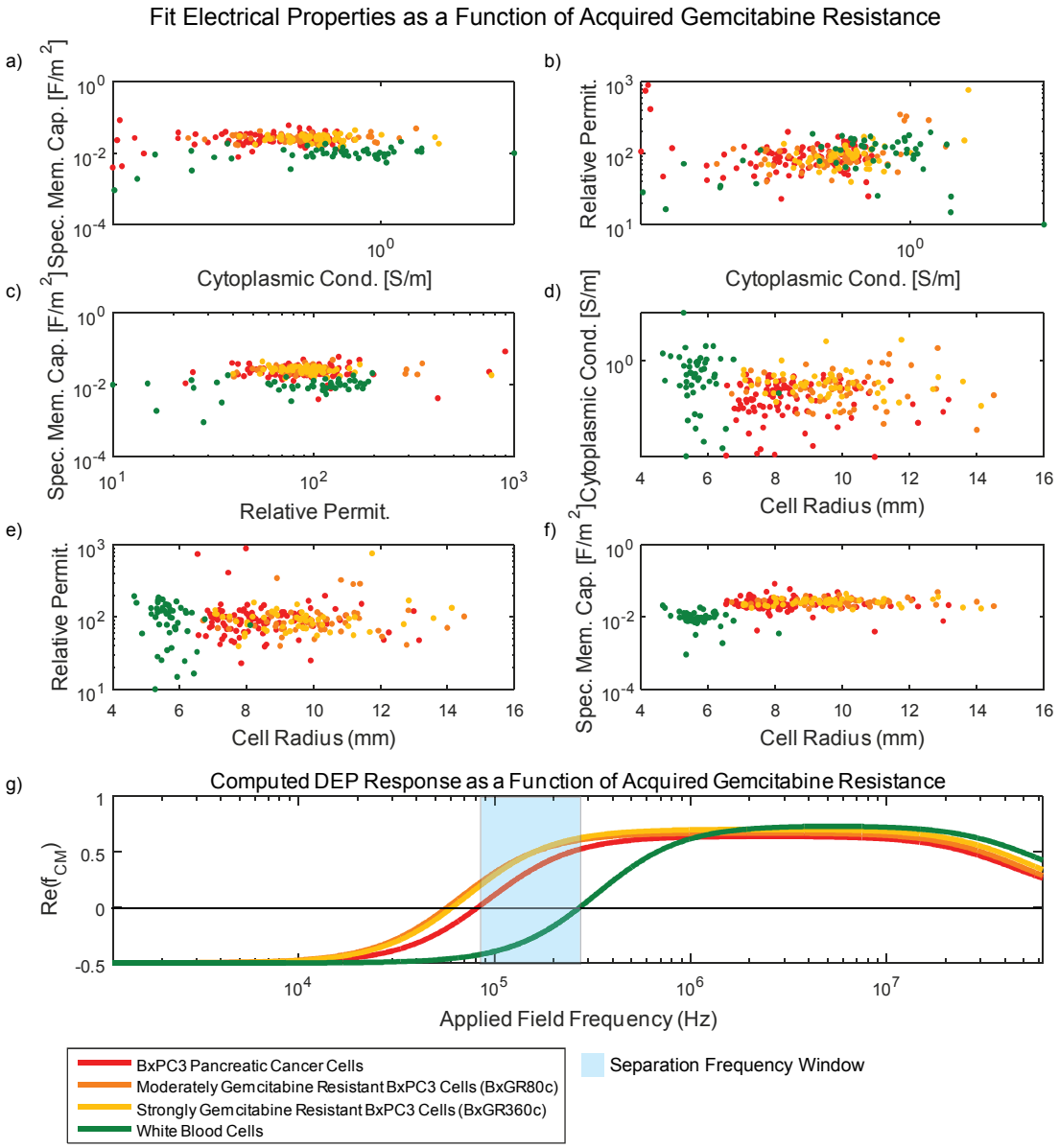

Figure 3.6 – a-f) Scatterplots of the electrical properties (fit from electrorotation spectra) for BxPC3s and BxGRs. Each point corresponds to the electrical properties obtained from an individual cell's electrorotation spectrum. g) Representative computed dielectrophoresis (DEP) spectra of BxPC3 pancreatic cancer cells and two gemcitabine-resistant BxPC3 sub-clones (a moderately resistant "BxGR 80c" and a strongly resistant "BxGR 360c") obtained by growing generations of BxPC3s in increasing concentrations of gemcitabine. Curves generated by taking the median of the cell electrical properties obtained from the cell type's electrorotation spectra fits. Note that the spectra for the two gemcitabine-resistant sub-clones nearly overlap. The analogously-obtained white blood cell fit properties (a-f) and computed DEP spectrum (g) are included for reference.

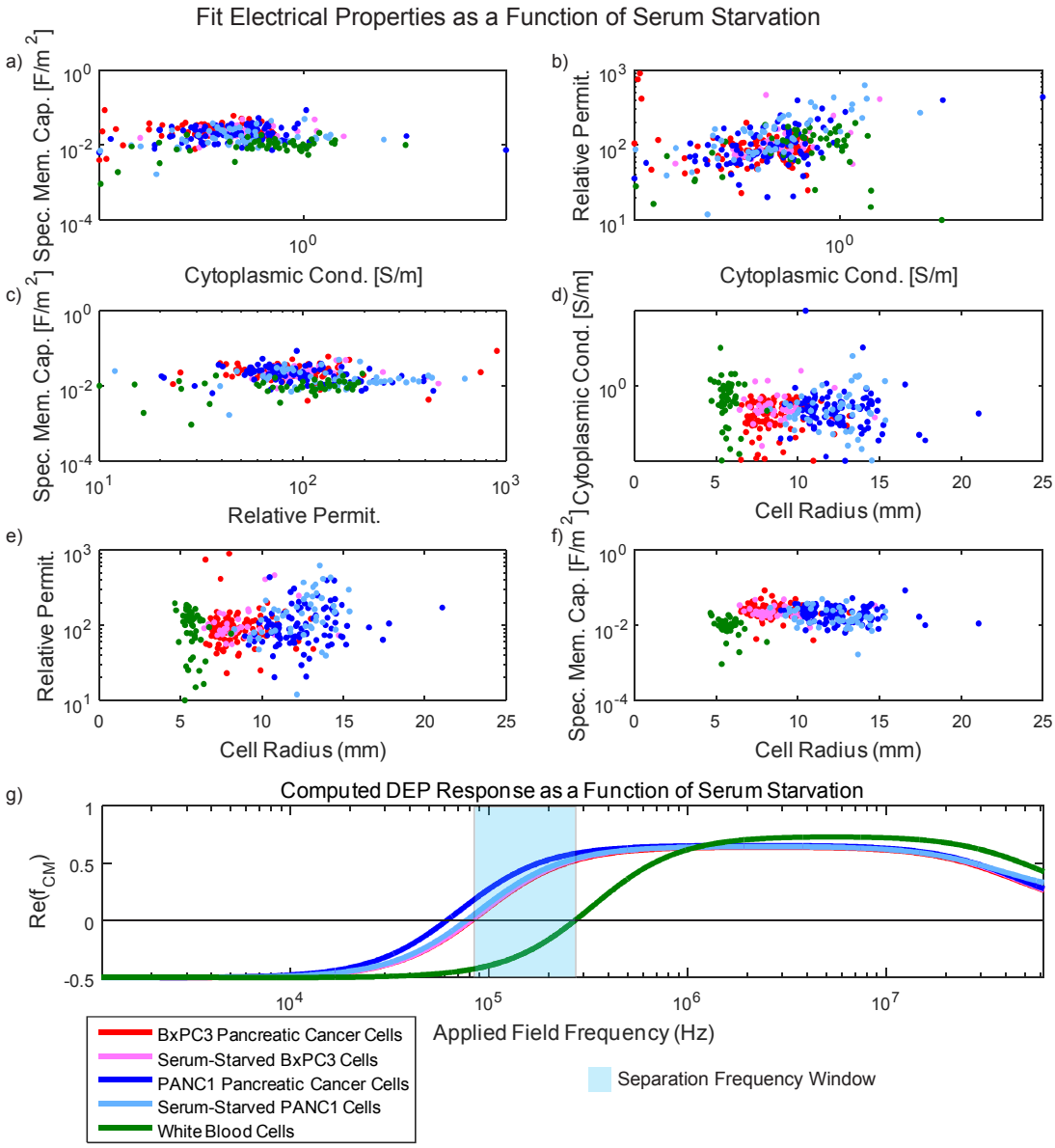

Figure 3.7 – a-f) Scatterplots of the electrical properties (fit from electrorotation spectra) for each cell type and serum condition. Each point corresponds to the electrical properties obtained from an individual cell's electrorotation spectrum. g) Representative computed dielectrophoresis (DEP) spectra of serum-starved and non-serum-starved pancreatic cancer cell lines (both BxPC3 and PANC1). Curves generated by taking the median of the cell electrical properties obtained from the cells' electrorotation spectra fits. Note that the spectrum for BxPC3s overlaps with the spectrum for serum-starved PANC1s. The analogously-obtained white blood cell fit properties (a-f) and computed DEP spectrum (g) are included for reference.

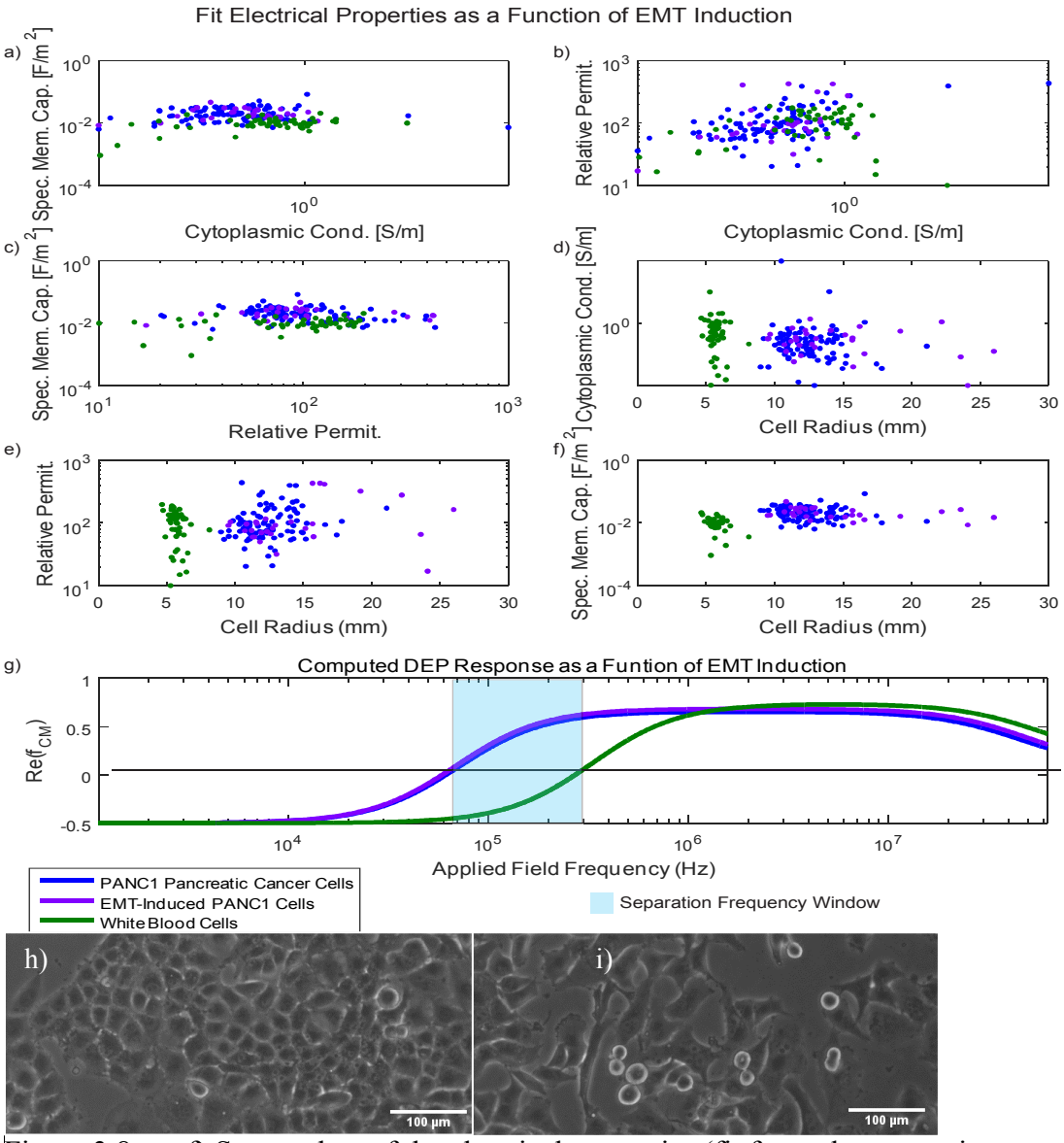

Figure 3.8 – a-f) Scatterplots of the electrical properties (fit from electrorotation spectra) for untreated and EMT-induced PANC1s. Each point corresponds to the electrical properties obtained from an individual cell's electrorotation spectrum. g) Representative DEP spectra of untreated PANC1 pancreatic cancer cells and PANC1s that have been treated with to induce the epithelial-to-mesenchymal transition (EMT). Curves generated by taking the median of the cell electrical properties obtained from the cells' electrorotation spectra fits. The analogously-obtained white blood cell fit properties (a-f) and computed DEP spectrum (g) are included for reference. h-i) Phase contrast images of the PANC1 cells (h) and the EMT-induced PANC1 cells (i), adhered to culture flask prior to liftoff.

Table 3.1 – Medians, (25-75 percentiles), and {95% bootstrapped confidence intervals} of electrical properties of white blood cells and cancer cell lines. Cytoplasmic conductivity, cytoplasmic relative permittivity, and membrane specific capacitance were computed by nonlinear curve fits to electrorotation spectra. Cell radii were measured by manually marking bright field images of cells. Crossover frequencies were computed for each cell according to its fit properties. Medium of conductivity was measured 70 mS/m and relative permittivity was calculated based on [30] to be 74.12. BxGR80c and BxGR360c are moderately- and strongly-gemcitabine resistant sub-clones (respectively) of BxPC3.

	Cytoplasmic Conductivity [mS/m]	Cytoplasmic Relative Permittivity [ ]	Membrane Specific Capacitance [mF/m^2]	Cell Radius [μm]	Calculated Crossover Frequency [kHz]	Significant* Difference in Crossover Freq. from WBC?	Significant* Difference in Crossover Freq. from Corresponding Untreated Cell?
White Blood Cells (N_{cells} = 49)	721 (530 - 977) {610 - 886}	111 (66 - 135) {79 - 122}	9.8 (8.1 - 11.9) { 9.0 - 11.0}	5.6 (5.4 - 5.9) { 5.5 - 5.8}	271 (224 - 334) {250 - 294}	Base Type	No Comparison Made
U937 (Lymphoma Cells) (N_{cells} = 57)	750 (606 - 903) {663 - 839}	106 (85 - 131) {95 - 118}	14.0 (11.4 - 17.3) {13.0 - 15.4}	7.5 (6.9 - 8.3) { 7.2 - 8.0}	147 (113 - 172) {132 - 163}	Yes	No Comparison Made
PANC1 (Pancreatic Cancer Cells) (N_{cells} = 97)	476 (328 - 607) {424 - 530}	90 (69 - 131) {77 - 101}	20.2 (14.4 - 28.6) {18.3 - 23.3}	12.1 (10.9 - 13.9) {11.7 - 12.6}	64 (44 - 80) { 56 - 67}	Yes	Base Type
Serum-Starved PANC1s (N_{cells} = 53)	462 (355 - 622) {401 - 530}	134 (81 - 216) {96 - 163}	15.4 (12.2 - 23.9) {13.2 - 18.1}	12.6 (10.7 - 13.6) {11.7 - 13.1}	81 (55 - 106) { 63 - 92}	Yes	Yes
EMT-Induced PANC1s (N_{cells} = 29)	536 (352 - 690) {442 - 647}	91 (66 - 107) {72 - 103}	20.1 (15.9 - 27.5) {16.8 - 24.9}	12.7 (11.5 - 15.8) {11.8 - 15.7}	54 (40 - 73) { 44 - 66}	Yes	No
BxPC3 (Pancreatic Cancer Cells) (N_{cells} = 101)	453 (320 - 571) {391 - 504}	91 (72 - 111) {84 - 98}	22.5 (19.6 - 28.7) {21.3 - 23.9}	8.1 (7.4 - 9.1) { 7.9 - 8.4}	80 (61 - 98) { 72 - 83}	Yes	Base Type
Serum-Starved BxPC3 Cells (N_{cells} = 39)	467 (417 - 568) {446 - 521}	88 (74 - 108) {76 - 94}	21.7 (18.8 - 26.1) {19.5 - 23.4}	8.5 (7.7 - 9.4) { 8.0 - 9.2}	82 (69 - 98) { 75 - 88}	Yes	No
BxGR80c, strongly-gemcitabine resistant subclone (N_{cells} = 55)	503 (395 - 608) {459 - 572}	85 (70 - 102) {81 - 94}	26.4 (22.6 - 29.6) {25.0 - 28.6}	9.9 (9.0 - 10.8) { 9.5 - 10.2}	56 (50 - 70) { 52 - 64}	Yes	Yes
BxGR360c, strongly-gemcitabine resistant subclone (N_{cells} = 54)	575 (470 - 673) {525 - 611}	92 (72 - 105) {81 - 99}	25.9 (22.6 - 31.0) {24.5 - 28.5}	9.6 (8.5 - 10.1) { 9.0 - 9.8}	59 (53 - 74) { 55 - 65}	Yes	Yes

*Median calculated crossover frequency of the cell type/treatment significantly differs ($p<0.05$, Wilcoxon rank sum test) from that of white blood cells (left column) or untreated cells of the same type (right column).

3.8 References

1. Danila DC, Heller G, Gignac G a, Gonzalez-Espinoza R, Anand A, Tanaka E, Lilja H, Schwartz L, Larson S, Fleisher M, Scher HI. Circulating tumor cell number and prognosis in progressive castration-resistant prostate cancer. Clin. Cancer Res. 2007;13:7053–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18056182. Accessed August 21, 2012.

2. Rhim AD, Mirek ET, Aiello NM, Maitra A, Bailey JM, McAllister F, Reichert M, Beatty GL, Rustgi AK, Vonderheide RH, Leach SD, Stanger BZ. EMT and dissemination precede pancreatic tumor formation. Cell 2012;148:349–61. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22265420. Accessed November 5, 2012.

3. De Bono JS, Scher HI, Montgomery RB, Parker C, Miller MC, Tissing H, Doyle G V, Terstappen LWWM, Pienta KJ, Raghavan D. Circulating tumor cells predict survival benefit from treatment in metastatic castration-resistant prostate cancer. Clin. Cancer Res. 2008;14:6302–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18829513. Accessed July 23, 2012.

4. Rhim AD, Thege FI, Santana SM, Lannin TB, Saha TN, Tsai S, Maggs LR, Kochman ML, Ginsberg GG, Lieb JG, Chandrasekhara V, Drebin J a., Ahmad N, Yang YX, Kirby BJ, Stanger BZ. Detection of circulating pancreas epithelial cells in patients with pancreatic cystic lesions. Gastroenterology 2014;146:647–651. Available at: http://dx.doi.org/10.1053/j.gastro.2013.12.007.

5. Kirby BJ, Jodari M, Loftus MS, Gakhar G, Pratt ED, Chanel-Vos C, Gleghorn JP, Santana SM, Liu H, Smith JP, Navarro VN, Tagawa ST, Bander NH, Nanus DM, Giannakakou P. Functional characterization of circulating tumor cells with a prostate-cancer-specific microfluidic device. PLoS One 2012;7:e35976. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3338784&tool=pmcentrez&rendertype=abstract. Accessed November 14, 2012.

6. Yu M, Ting DT, Stott SL, Wittner BS, Ozsolak F, Ciciliano JC, Smas ME, Winokur D, Gilman AJ, Ulman MJ, Xega K, Contino G, Alagesan B, Brian W, Milos PM, Ryan DP, Sequist L V, Bardeesy N, Toner M, Maheswaran S, Haber D a, Paul S, Ciciliano JC, Smas ME, Winokur D, Gilman AJ, Ulman MJ, Xega K, Contino G, Alagesan B, Brannigan BW, Milos PM, Ryan DP, Sequist L V, Bardeesy N, Ramaswamy S, Toner M, Maheswaran S, Haber D a. RNA sequencing of pancreatic circulating tumour cells implicates WNT signalling in metastasis. Nature 2013;487:510–3.

7. Lin HK, Zheng S, Williams AJ, Balic M, Groshen S, Scher HI, Fleisher M, Stadler W, Datar RH, Tai Y-C, Cote RJ. Portable filter-based microdevice for detection and characterization of circulating tumor cells. Clin. Cancer Res. 2010;16:5011–8. Available at:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2955786&tool=pmcentrez&rendertype=abstract. Accessed January 14, 2014.

8. Shim S, Gascoyne P, Noshari J, Hale KS. Dynamic physical properties of dissociated tumor cells revealed by dielectrophoretic field-flow fractionation. Integr. Biol. (Camb). 2011;3:850–62. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21691666. Accessed May 14, 2013.

9. Allard WJ, Matera J, Miller MC, Repollet M, Connelly MC, Rao C, Tibbe AGJ, Uhr JW, Terstappen LWMM. Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases. Clin. Cancer Res. 2004;10:6897–904. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15501967. Accessed July 15, 2012.

10. Gleghorn JP, Pratt ED, Denning D, Liu H, Bander NH, Tagawa ST, Nanus DM, Giannakakou P a, Kirby BJ. Capture of circulating tumor cells from whole blood of prostate cancer patients using geometrically enhanced differential immunocapture (GEDI) and a prostate-specific antibody. Lab Chip 2010;10:27–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3031459&tool=pmcentrez&rendertype=abstract. Accessed August 15, 2013.

11. Nagrath S, Sequist L V, Maheswaran S, Bell DW, Irimia D, Ulkus L, Smith MR, Kwak EL, Digumarthy S, Muzikansky A, Ryan P, Balis UJ, Tompkins RG, Haber D a, Toner M. Isolation of rare circulating tumour cells in cancer patients by microchip technology. Nature 2007;450:1235–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3090667&tool=pmcentrez&rendertype=abstract. Accessed November 13, 2013.

12. Gorges TM, Tinhofer I, Drosch M, Röse L, Zollner TM, Krahn T, von Ahsen O. Circulating tumour cells escape from EpCAM-based detection due to epithelial-to-mesenchymal transition. BMC Cancer 2012;12:178. Available at: http://www.biomedcentral.com/1471-2407/12/178.

13. Wang Z, Li Y, Kong D, Banerjee S, Ahmad A, Azmi AS, Ali S, Abbruzzese JL, Gallick GE, Sarkar FH. Acquisition of epithelial-mesenchymal transition phenotype of gemcitabine-resistant pancreatic cancer cells is linked with activation of the notch signaling pathway. Cancer Res. 2009;69:2400–7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2657919&tool=pmcentrez&rendertype=abstract. Accessed November 14, 2014.

14. Yang J, Huang Y, Wang XB, Becker FF, Gascoyne PR. Cell separation on microfabricated electrodes using dielectrophoretic/gravitational field-flow fractionation. Anal. Chem. 1999;71:911–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10079757.

15. Shim S, Stemke-Hale K, Tsimberidou AM, Noshari J, Anderson TE, Gascoyne PRC. Antibody-independent isolation of circulating tumor cells by continuous-flow dielectrophoresis. Biomicrofluidics 2013;7:011807. Available at: http://link.aip.org/link/BIOMGB/v7/i1/p011807/s1&Agg=doi. Accessed December 19, 2013.

16. Becker FF, Wang XB, Huang Y, Pethig R, Vykoukal J, Gascoyne PR. Separation of human breast cancer cells from blood by differential dielectric affinity. Proc. Natl. Acad. Sci. U. S. A. 1995;92:860–4. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=42720&tool=pmcentrez&rendertype=abstract.

17. Huang C, Smith JP, Saha TN, Rhim AD, Kirby BJ. Characterization of microfluidic shear-dependent epithelial cell adhesion molecule immunocapture and enrichment of pancreatic cancer cells from blood cells with dielectrophoresis. Biomicrofluidics 2014;8:044107. Available at: http://scitation.aip.org/content/aip/journal/bmf/8/4/10.1063/1.4890466.

18. Huang C, Santana SM, Liu H, Bander NH, Hawkins BG, Kirby BJ. Characterization of a hybrid dielectrophoresis and immunocapture microfluidic system for cancer cell capture. Electrophoresis 2013:2970–2979. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23925921. Accessed November 13, 2013.

19. Huang C, Liu H, Bander NH, Kirby BJ. Enrichment of prostate cancer cells from blood cells with a hybrid dielectrophoresis and immunocapture microfluidic system. Biomed. Microdevices 2013;15:941–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23807279. Accessed January 14, 2014.

20. Smith JP, Huang C, Kirby BJ. Enhancing sensitivity and specificity in rare cell capture microdevices with dielectrophoresis. Biomicrofluidics 2015;9:1–14.

21. Hawkins BG, Huang C, Arasanipalai S, Kirby BJ. Automated Dielectrophoretic Characterization of Mycobacterium smegmatis. Anal. Chem. 2011;83:3507–3515. Available at: http://pubs.acs.org/doi/abs/10.1021/ac2002017.

22. Morgan H, Green NG. AC Electrokinetics: Colloids and Nanoparticles. Hertfordshire, England: Research Studies Press Ltd.; 2003.

23. Kirby BJ. Micro- and nanoscale fluid mechanics : Transport in Microfluidic Devices. 1st ed. Cambridge: Cambridge University Press; 2010. Available at: http://www.kirbyresearch.com/index.cfm/wrap/textbook/microfluidicsnanofluidics.html.

24. Jones TB. Electromechanics of Particles. Cambridge: Cambridge University Press; 1995. Available at: http://dx.doi.org/10.1017/CBO9780511574498. Accessed February 18, 2014.

25. Ratanachoo K, Gascoyne PRC, Ruchirawat M. Detection of cellular responses to toxicants by dielectrophoresis. Biochim. Biophys. Acta 2002;1564:449–58. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2726261&tool=pmcentrez&rendertype=abstract.

26. Thege FI, Lannin TB, Saha TN, Tsai S, Kochman ML, Hollingsworth M a, Rhim AD, Kirby BJ. Microfluidic immunocapture of circulating pancreatic cells using parallel EpCAM and MUC1 capture: characterization, optimization and downstream analysis. Lab Chip 2014;14:1775–84. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24681997. Accessed October 4, 2014.

27. Shah AN, Summy JM, Zhang J, Park SI, Parikh NU, Gallick GE. Development and characterization of gemcitabine-resistant pancreatic tumor cells. Ann. Surg. Oncol. 2007;14:3629–37. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17909916. Accessed November 18, 2014.

28. Park S, Ang RR, Duffy SP, Bazov J, Chi KN, Black PC, Ma H. Morphological Differences between Circulating Tumor Cells from Prostate Cancer Patients and Cultured Prostate Cancer Cells. PLoS One 2014;9:e85264. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24416373. Accessed January 16, 2014.

29. Rao C, Chianese D, Doyle G, Miller M, Russell T, Sanders R, Terstappen L. Expression of epithelial cell adhesion molecule in carcinoma cells present in blood and primary and metastatic tumors. Int. J. Oncol. 2005. Available at: http://www.spandidos-publications.com/ijo/27/1/49.

30. Malmberg CG, Maryott a. a. Dielectric constants of aqueous solutions of dextrose and sucrose. J. Res. Natl. Bur. Stand. (1934). 1950;45:299.

31. Huang Y, Pethig R. Electrode design for negative dielectrophoresis. Meas. Sci. Technol 1991;1142:2–7.

32. Archer S, Morgan H, Rixon FJ. Electrorotation studies of baby hamster kidney fibroblasts infected with herpes simplex virus type 1. Biophys. J. 1999;76:2833–42. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1300254&tool=pmcentrez&rendertype=abstract.

33. Rohani A, Varhue W, Su Y, Swami NS. Electrical tweezer for highly parallelized electrorotation measurements over a wide frequency bandwidth. Electrophoresis 2014;35:1795–802. Available at: http://doi.wiley.com/10.1002/elps.201400021\nhttp://www.ncbi.nlm.nih.gov/pubmed/24668830.

34. Gasperis G De, Wang X, Yang J, Becker FF, Gascoyne PRC. Automated electrorotation: dielectric characterization of living cells by real-time motion estimation. Meas. Sci. Technol. 1998;9:518–529. Available at: http://stacks.iop.org/0957-0233/9/i=3/a=029?key=crossref.61bafdcb1d8b9f96eb7f186589639926.

35. Wilson C a., Theriot J a. A correlation-based approach to calculate rotation and translation of moving cells. IEEE Trans. Image Process. 2006;15:1939–1951.

36. Qihua Y, Hao L, Ganghai L, Yanling L. Cell Electrorotation Motion Parameters Detection Based on Image Processing. 2009 WRI World Congr. Comput. Sci. Inf. Eng. 2009:549–553. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5170760. Accessed July 17, 2013.

37. Gascoyne PRC, Jody Vykoukal. Particle separation by dielectrophoresis. Electrophoresis 2002;23:1973–1983.

38. Čížek P, Víšek JÁ. Least Trimmed Squares. In: XploRe® - Appl. Guid. Berlin, Heidelberg: Springer Berlin Heidelberg; 2000. p 49–63. Available at: http://link.springer.com/10.1007/978-3-642-57292-0_2.

39. Lazar DC, Cho EH, Luttgen MS, Metzner TJ, Uson ML, Torrey M, Gross ME, Kuhn P. Cytometric comparisons between circulating tumor cells from prostate cancer patients and the prostate-tumor-derived LNCaP cell line. Phys. Biol. 2012;9:016002. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3387997&tool=pmcentrez&rendertype=abstract. Accessed July 16, 2012.

40. Antonyak MA, Li B, Lindsey K, Johnson JL, Druso JE, Bryant KL, Holowka DA, Cerione RA, Boroughs LK. Cancer cell-derived microvesicles induce transformation by transferring tissue transglutaminase and fi bronectin to recipient cells. 2011;108.

41. Weiss L, Ward PM. Cell detachment and metastasis. Cancer Metastasis Rev. 1983;2:111–127.

42. Buonato JM, Lan IS, Lazzara MJ. EGF augments TGF -induced epithelial-mesenchymal transition by promoting SHP2 binding to GAB1. J. Cell Sci. 2015;128:3898–3909. Available at: http://jcs.biologists.org/cgi/doi/10.1242/jcs.169599.

43. Wang X, Becker FF, Gascoyne PRC. The fractal dimension of cell membrane correlates with its capacitance: a new fractal single-shell model. Chaos 2010;20:043133. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3017572&tool=pmcentrez&rendertype=abstract. Accessed September 16, 2013.

44. Gascoyne PRC, Shim S, Noshari J, Becker FF, Stemke-Hale K. Correlations between the dielectric properties and exterior morphology of cells revealed by dielectrophoretic field-flow fractionation. Electrophoresis 2013;34:1042–50. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23172680. Accessed December 19, 2013.

45. Salmanzadeh A, Sano MB, Gallo-Villanueva RC, Roberts PC, Schmelz EM, Davalos R V. Investigating dielectric properties of different stages of syngeneic murine ovarian cancer cells. Biomicrofluidics 2013;7:11809. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24403991. Accessed January 22, 2014.

46. Mulhall HJ, Labeed FH, Kazmi B, Costea DE, Hughes MP, Lewis MP. Cancer, pre-cancer and normal oral cells distinguished by dielectrophoresis. Anal. Bioanal. Chem. 2011;401:2455–2463.

47. Sano MB, Henslee E a, Schmelz E, Davalos R V. Contactless dielectrophoretic spectroscopy: examination of the dielectric properties of cells found in blood. Electrophoresis 2011;32:3164–71. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22102497. Accessed March 11, 2014.

CHAPTER 4

Cytoplasmic permittivity and conductivity computed from electrorotation spectra negatively correlate with BODIPY fluorescence in nitrogen-starved *Chlamydomonas reinhardtii.*[3]

4.1 Abstract

Algae are promising feedstocks for biofuels, and there is a critical need for a rapid, inexpensive, and label-free measurement of lipid accumulation in algae cells. Measuring the electrical properties of algae has shown promise for monitoring lipid accumulation because lipid accumulation correlates with a decrease in effective cytoplasmic conductivity. Previous models, however, have often assumed a constant cytoplasmic permittivity through the lipid accumulation process. We have used automated electrorotation to measure electrical properties of *Chlamydomonas reinhardtii* cells undergoing lipid accumulation. We have also measured an increase in BODIPY fluorescence measured via flow cytometry for the nitrogen-starved *Chlamydomonas reinhardtii*. Median BODIPY fluorescence significantly correlated ($p<0.05$, F-test) with fit electrorotation spectrum parameters, indicating that cytoplasmic conductivity and permittivity are both negatively correlated with lipid accumulation.

[3] Parts of this chapter have been submitted for publication: Cytoplasmic permittivity and conductivity computed from electrorotation spectra negatively correlate with BODIPY fluorescence in nitrogen-starved *Chlamydomonas reinhardtii*. Timothy Lannin, Wey-Wey Su, and Brian Kirby. *Bioresource Technology*.

4.2 Motivation

In a world of increasingly limited non-renewable resources and ever-increasing energy demands, suitable alternative renewable resources are becoming more and more necessary. Among options for renewable liquid fuels, algae have incredible potential as a sustainable source due to their fast growth rate, fixation of carbon from the atmosphere, and ability to thrive even in wastewater (1,2). Algae are further promising because they are incredibly productive, yielding much more oil per hectare than other plant options, including corn and soybeans (1).

For energy and fuel applications, the extraction of lipids from algae are of interest because hydrocarbon chains of lipid molecules can be easily processed into biodiesel through transesterification (3,4), which can fuel regular diesel engines. Furthermore, it has been shown that environmental stresses can induce algal lipid accumulation (1,5), and that in particular, nitrogen starvation of several algae species has been immensely successful (1,6), increasing lipids up to three times as much (7) in some species. Unfortunately, this increase comes at the cost of halting overall algae cell growth, and selecting the optimal harvest time is necessary in order to take advantage of the lipid increase and maximize biodiesel yield from the algal cultures (6,8,9). Additionally, natural variations in sunlight and or temperature make it difficult to predict the lipid accumulation of algae accurately (10,11). There is therefore a need for real-time lipid measurement to indicate maximum lipid content in order to inform optimal harvest times.

Traditional methods of measuring lipid content have drawbacks. Direct measurements such as gravimetric determination usually have multiple steps that take

hours to days, and require large samples (12) or extremely accurate mass measurements (13), and fractional losses vary depending on extraction method (14). High performance liquid chromatography (HPLC) is also a method requiring smaller sample sizes that has been successfully used to analyze lipids in algae (15), but the technique also is labor intensive and can be potentially confounded by chlorophyll extracted alongside the lipids (16). Another technique, Fourier Transform Infrared (FTIR) spectroscopy, slows data acquisition by requiring sample drying, and staining protocols with Nile Red require precise dye concentration and measurement (17,18), requiring protocol optimization. Additionally, fluorescent dye uptake can be variable (19). Finally, nuclear magnetic resonance (NMR) spectroscopy can yield accurate, quantitative lipid measurements nondestructively (20), but NMR spectrometer cost and operation difficulty prevents widespread use.

Electrokinetic techniques such as dielectrophoresis (DEP) or impedance cytometry are suitable because these methods can quickly return information about dielectric properties (related to lipid content) of cells nondestructively. Previous research has shown that upper crossover frequency, a property linked to the cytoplasmic permittivity and conductivity (21–25), is different for high- and low-lipid algae cells (26,27). Bulk electrical impedance measurements on *Chlamydomonas reinhardtii* connect a shift in scattering parameter critical frequency, consistent with a change in cytoplasmic conductivity, to increased lipid content(28). Impedance cytometry also quickly measures differences in dielectric properties between algae cells of high and low lipid content (20). Finally, DEP devices have been fabricated that nondestructively sort cells with a higher lipid content from a population (29–31).

Although DEP and impedance cytometry are well-suited for sorting and quick measurements, respectively, there are drawbacks to using these two methods for studying cells. As shown by (32) and discussed in the context of algae by (26,27), DEP upper crossover frequency depends on both cytoplasmic conductivity and permittivity, so an assumption of known cytoplasmic permittivity is necessary to estimate conductivity from a single crossover frequency measurement. However, changes in lipid content in algae cells likely change both conductivity and permittivity because bulk lipids have both a lower conductivity and permittivity than water. Impedance cytometry has been used to measure a dense spectrum, sacrificing single-cell measurements for spectral resolution, or a sparse single-cell spectrum, which require assumptions of cell parameters. In a configuration to yield average cell properties, impedance cytometry measurements indicated that both cytoplasmic permittivity and conductivity decreased with lipid accumulation (20). With this knowledge, we seek to understand know how both conductivity and permittivity change on a single-cell basis, which necessitates acquiring single-cell, instead of population-averaged, dielectric spectra.

DEP spectrometry has been presented in the literature, but it is only suited to acquiring information in the regime of positive DEP, yielding incomplete spectra (33). A related technique, electrorotation (ROT) (21–23), measures the external-field-frequency-dependent rotation rates of cells from torque inducted by a rotating electric field. Compared to DEP, in which response depends strongly on the position of the cells in the device, ROT may be created experimentally with approximately uniform

fields (34,35), which reduces modeling complexity. ROT spectra may be related to DEP spectra via the Kramers-Kronig relations (21).

Electrorotation, through our experimental methods, can inform measurement of many cellular properties simultaneously. Electrorotation has been used to characterize snow algae (36) as well as *Chlorella protothecoides* in different nutrient conditions (37). Electrorotation spectra measurements afford a thorough characterization of the electrical property changes with increased lipid content, which inform design of effective DEP cell sorters and impedance cytometry measurement tools. In this study, we measure electrorotation spectra of both nitrogen-replete and nitrogen-starved *Chlamydomonas reinhardtii*, and extract single-cell electrical properties, that will allow us to correlate these fit ROT parameters to amounts of lipid accumulation.

4.3 Theory

AC Electrokinetic theory (22,23,25,38) can describe approximate ROT peak magnitudes and locations. Typically, conductivity and permittivity are combined into a frequency-dependent complex permittivity:

$$\underline{\varepsilon} = \varepsilon + \frac{\sigma}{j\omega} \tag{1}$$

where ε is the sphere permittivity, σ is the conductivity, j is the imaginary number, and ω is the angular frequency of the field. For concentric, multi-shelled spheres, the effective permittivity of the sphere may be written in terms of its components

$$\underline{\varepsilon_{eff}} = \underline{\varepsilon_2}\left[\frac{\frac{a^3}{(a-\delta)^3} + 2\frac{\underline{\varepsilon_1} - \underline{\varepsilon_2}}{\underline{\varepsilon_1} + 2\underline{\varepsilon_2}}}{\frac{a^3}{(a-\delta)^3} - \frac{\underline{\varepsilon_1} - \underline{\varepsilon_2}}{\underline{\varepsilon_1} + 2\underline{\varepsilon_2}}}\right] \tag{2}$$

where a is the outer radius, $(a-\delta)$ is the inner radius, subscript 2 is the shell and subscript 1 is the core. This formula may be called recursively for multiple shells (e.g. a cell wall surrounding a cell membrane surrounding cytoplasm). In the limit where the shell thickness is small compared to the radius and the shell permittivity and conductivity are small compared to those of the core, the complex permittivity reduces to

$$\underline{\varepsilon_{eff}} = \frac{\underline{\varepsilon_1}\,\underline{\varepsilon_2}}{\frac{\delta}{a}\underline{\varepsilon_1} + \underline{\varepsilon_2}} = \frac{\underline{\varepsilon_1}\; a\,(C'' + \frac{G''}{j\omega})}{\underline{\varepsilon_1} + a\,(C'' + \frac{G''}{j\omega})} \tag{3}$$

where δ is the shell thickness, $C''$is the specific capacitance of the shell (SI units if F/m^2), and $G''$is the specific conductance of the shell (SI units of S/m^2). The dipole induced when the sphere is placed in an externally-applied uniform electric field depends on the complex permittivity:

$$\underline{\vec{\boldsymbol{p}}_{\mathbf{0}}} = 4\pi\,\varepsilon_m \vec{\boldsymbol{E}}_{\mathbf{0}}\,a^3\,\frac{\underline{\varepsilon_p} - \underline{\varepsilon_m}}{\underline{\varepsilon_p} + 2\underline{\varepsilon_m}} = 4\pi\,\varepsilon_m \vec{\boldsymbol{E}}_{\mathbf{0}}\,R^3\,\underline{f_{CM}} \tag{4}$$

where ε_m is the complex permittivity of the suspending medium, and $\underline{f_{CM}}$ is the complex Clausius-Mossotti factor. If the external electric field is nonuniform, the dipole interacts with the field gradient to yield a time-averaged dielectrophoretic force:

$$\langle \vec{\boldsymbol{F}}_{DEP} \rangle = \pi\,\varepsilon_m\,a^3\,\mathrm{Re}\left(\underline{f_{CM}}\right)\,\nabla(\vec{\boldsymbol{E}}_{\mathbf{0}} \cdot \vec{\boldsymbol{E}}_{\mathbf{0}}) \tag{5}$$

If the external electric field is constant in magnitude, but rotating as a function of time the dipole-field interaction yields an electrorotation torque:

$$\langle \vec{\boldsymbol{T}}_{ROT} \rangle = -4\pi\,\varepsilon_m\,a^3\,\mathrm{Im}\left(\underline{f_{CM}}\right)\left[\mathrm{Re}\left(\underline{\vec{\boldsymbol{E}}_{\mathbf{0}}}\right) \times \mathrm{Im}\left(\underline{\vec{\boldsymbol{E}}_{\mathbf{0}}}\right)\right] \tag{6}$$

Thus, the frequency-dependent complex Clausius-Mossotti factor is the key parameter affecting dielectrophoresis (real part) and electrorotation (imaginary part). The imaginary part of the Clausius-Mossotti factor is proportional to the slope of the real part through the Kramers-Kronig relation; peaks in ROT spectra correspond to inflection points in DEP spectra. In the inertia-free limit of Stokes flow, viscous torque balances ROT torque, leading to a field-frequency-dependent steady rotation rate of cells proportional to Im $\left(\underline{f_{CM}}\right)$. In most experimental electrorotation spectra of cells, two clear peaks are present: a low-frequency peak around 10 kHz and a high-frequency peak around 20 MHz. Because these peaks are widely-spaced, an approximate expression for frequency-dependent ROT rate can be written as

$$\Omega(\omega) = B \text{ Im}\left(\underline{f_{CM}}\right) = \frac{\Omega_1 \, \omega \, \tau_1}{(\omega\tau_1)^2 + 1} + \frac{\Omega_2 \, \omega \, \tau_2}{(\omega\tau_2)^2 + 1} \tag{7}$$

where B is a proportionality constant accounting for electric field strength, (real) permittivity of the suspending medium, and friction. The peak locations occur at angular field frequencies of $1/\tau_1$ and $1/\tau_2$ and have approximate peak amplitudes of $\Omega_1/2$ and $\Omega_2/2$. The algebra connecting ROT peak parameters $(\tau_1, \tau_2, \Omega_1, \Omega_2)$ with medium (subscript m), cytoplasm (subscript c), and shell physical and electrical properties $(R, \varepsilon_m, \sigma_m, \varepsilon_c, \sigma_c, C'', G'')$ is complicated, but may be simplified when 1) the effects of the wall and membrane may be lumped together into a single shell, 2) the conductivity of the cytoplasm dominates over the conductivity of the suspending medium and the trans-shell conductance, 3) the first peak occurs at a low enough frequency to neglect effects of permittivity (except for shell specific capacitance), and 4) the second peak occurs at a high enough frequency to short out the shell capacitance and neglect its effect. Thus, the two approximate peak parameters can be derived

separately from one another. Considering the first peak with assumptions 1) and 3), the complex permittivity of the particle is:

$$\underline{\varepsilon_{low\ frequency}} = \frac{\frac{\sigma_c}{j\omega}\, a\, (C'' + \frac{G''}{j\omega})}{\frac{\sigma_c}{j\omega} + a\, (C'' + \frac{G''}{j\omega})} \tag{8}$$

Putting this expression into the expression for Clausius-Mossotti factor, also with assumption 3), taking the imaginary part, and applying assumption 2), we arrive at:

$$\mathrm{Im}\left(\underline{f_{CM\ low\ frequency}}\right) = \frac{\frac{3\, a\, C''\sigma_m}{(2\sigma_m + a\, G'')^2}\, \omega}{\omega^2 \left(\frac{a\, C''}{2\sigma_m + a\, G''}\right)^2 + 1} \tag{9}$$

Matching coefficients of ω, we get an expression for the first peak parameters:

$$\tau_1 = \frac{a\, C''}{2\sigma_m + a\, G''} \tag{10}$$

$$\Omega_1 = B\frac{3\sigma_m}{2\sigma_m + a\, G''} \tag{11}$$

Considering the second peak with assumption 4), the Clausius-Mossotti factor is that of a homogenous sphere:

$$\underline{f_{CM\ high\ frequency}} = \frac{(\sigma_c - \sigma_m) + j\omega(\varepsilon_c - \varepsilon_m)}{(\sigma_c + 2\sigma_m) + j\omega(\varepsilon_c + 2\varepsilon_m)} \tag{12}$$

Multiplying numerator and denominator by the complex conjugate of the denominator and taking the imaginary part, we arrive at:

$$\mathrm{Im}\left(\underline{f_{CM\ high\ frequency}}\right) = \frac{\left(\frac{3\sigma_m}{\sigma_c + 2\sigma_m} - \frac{3\varepsilon_m}{\varepsilon_c + 2\varepsilon_m}\right)\left(\frac{\varepsilon_c + 2\varepsilon_m}{\sigma_c + 2\sigma_m}\right)\, \omega}{\omega^2 \left(\frac{\varepsilon_c + 2\varepsilon_m}{\sigma_c + 2\sigma_m}\right)^2 + 1} \tag{13}$$

Matching coefficients of ω and then simplifying with assumption 2), we get the well-known Maxwell-Wagner time constant associated with the second peak location:

$$\tau_2 = \tau_{MW} = \frac{\varepsilon_c + 2\varepsilon_m}{\sigma_c} \tag{14}$$

$$\Omega_2 = B\frac{-3\varepsilon_m}{\varepsilon_c + 2\varepsilon_m} \tag{15}$$

These simplified expressions are consistent with those in (24). Because the ROT-rate-to-Clausius-Mossotti factor scale, B, is notoriously difficult to model even in simple systems such as liposomes (39), a measurement of an electrorotation spectrum that yields τ_1, τ_2, Ω_1, and Ω_2 from a curve fit, along with external measurements of $a, \varepsilon_m, \sigma_m$, yields four equations ((10)(11)(14)(15)) for five unknowns $(B, \varepsilon_c, \sigma_c, C'', G'')$. Thus, it is most appropriate to consider ratios of measured quantities that are proportional to electrical properties of interest, because it is impossible to solve for all unknown properties directly:

$$-\frac{\Omega_1}{\Omega_2\,\tau_2} = \frac{\sigma_c}{2\varepsilon_m(1 + \frac{a\,G''}{2\sigma_m})} \tag{16}$$

$$-\frac{\Omega_1}{\tau_2\,\tau_1\,\Omega_2} = \frac{\sigma_c\,\sigma_m}{\varepsilon_m a\,C''} \tag{17}$$

$$-\frac{\Omega_1}{\Omega_2} = \frac{\varepsilon_c + 2\varepsilon_m}{2\varepsilon_m\left(1 + \frac{a\,G''}{2\sigma_m}\right)} \tag{18}$$

$$-\frac{\Omega_1}{\tau_1\Omega_2} = \frac{(\varepsilon_c + 2\varepsilon_m)\,\sigma_m}{\varepsilon_m\,a\,C''} \tag{19}$$

For these relaxation ratios, the absolute magnitude of ROT peaks is unimportant, only the locations and relative amplitude. Note that measured Ω_1 and Ω_2 have opposite sign. The qualitative effect of changes in $\varepsilon_c, \sigma_c, C''$and G'' can be seen in Figure 4.1, which shows simulations of ROT spectra without making approximations 2), 3) and 4).

4.4 Materials and Methods

4.4.1 Algae Culture

Chlamydamonas reinhardtii (CC-125, mt+) purchased from UTEX culture collection grew in 20 mL Pyrex culture tubes with loosened caps to allow for gas exchange. TAP media composed of deionized water with 10 mL 5x concentrated Beijerinck's solution (750mM NH_4Cl, 35mM $CaCl_2$, 40mM $MgSO_4 \cdot 7H_2O$), 8.33 mL phosphate solution (98mM K_2HPO_4,54mM KH_2PO_4), 10 mL Tris-Acetate buffer (2.00M tris (hydroxymethyl) aminomethane, 1.75M acetic acid), and 1 mL Hunter's trace metal solution (134mM $Na_2EDTA \cdot 2H_2O$, 77mM $ZnSO_4 \cdot 7H_2O$, 184mM H_3BO_3, 26mM $MnCl_2 \cdot 4H_2O$, 18mM $FeSO_4 \cdot 7H_2O$, 7mM $CoCl_2 \cdot 6H_2O$, 5mM $CuSO_4 \cdot 5H_2O$, 8 μM $(NH_4)_6Mo_7O_{24} \cdot 4H_2O$) per 1 L of media. Beijerinck's solution was prepared with two variants: one nitrogen-replete version described above and another using NaCl to replace NH_4Cl in order to create a nitrogen-limited variant. These recipes were used by (28), from (40) and (41). Five mL of TAP media were aliquoted into the culture tubes and then autoclaved and stored at 4C. Passaging occurred twice per week and consisted of inoculating 150 μL parent culture into a TAP aliquot.

We grew algae cultures in a constant illumination growth chamber with adjustable LED light panels (color temperature 5500 K) from Fancier Photographic Equipment (Ningbo, China) adjusted to deliver 100 ± 20 μmol / m^2s photon flux density light to the cultures. In addition to biweekly passaging, cultures were gently agitated once per day via Vortex Genie 1 orbital shaker (Scientific Industries, New York).

For nitrogen starvation, regular algae cultures were grown in nitrogen-replete media for three days before being split into 2 mL aliquots and then centrifuged at 3000g for 14 minutes. We then resuspended one cell pellet in a nitrogen-limited culture tube and one in a nitrogen-replete tube to serve as a control. We conducted two sets of electrorotation experiments, one on days 1 and 3 and another on days 3 and 5 after the induction of nitrogen starvation.

4.4.2 Cell Preparation for ROT

We prepared a low-conductivity isotonic sugar buffer with 19.81 g/L sucrose, 2 g/L Pluronic F68 from Sigma-Aldrich (St. Louis, MO) in deionized water, and added TAP media to attain a medium conductivity of 10.6 mS/m. This buffer was aliquoted into smaller flasks and then autoclaved in order to use as the medium for electrorotation. The buffer was refrigerated at 4C and warmed up to room temperatures before experimentation. Relative permittivity was calculated based on (42) to be 74.2.

We prepared electrorotation algae samples by taking 400 µL of algae culture, adding 24µL of 1M acetic acid and gently vortexing for 45 seconds before neutralizing the acid with 24 µL of 1M NaOH. The quick pH treatment immobilized the algae cells by removing flagella (43), allowing us to take electrorotation measurements of the otherwise extremely motile cultures. Following the treatment, we washed three times by centrifuging the cells at 3500xg for 7 minutes and resuspending the pellet in ROT buffer. The final resuspension was diluted additionally in ROT buffer until cells were sufficiently well-spaced in the ROT device. Video capture and ArbStudio frequency

sweeps were started once enough cells had settled to the surface. Between running samples, we rinsed the well thoroughly at least three times with deionized water before resuming.

4.4.3 Experimental Apparatus and Application of ROT Fields

A 4-channel function generator (ArbStudio 1104, Teledyne Lecroy) energized an electrorotation device with a hyperbolic 4-electrode geometry (34) with 3.8 $V_{p\text{-}p}$ across a 800 μm tip-to-tip spacing. It applied 36 frequencies, log-spaced from 1.125 kHz-62.5 MHz for 5-second windows in a shuffled order. An Eclipse TE2000U inverted microscope by Nikon (Melville, NY) equipped with a 40x/0.6 Plan Fluor objective, a RETIGA EXi FAST, and NIS Elements D software captured ROT videos under brightfield illumination. All other details on the experimental apparatus and applications of ROT fields are the same as those in (44).

4.4.4 Video Processing and Curve Fitting

An automated MATLAB script enabled extraction of ROT rates from captured ROT videos. This process and the subsequent curve fitting are described in (44). Algae video processing was identical except for the background subtraction protocol, which was omitted due to artifacts associated with high contrast of algae cells. Curve fitting used the same least trimmed squares strategy, but fit $\langle \ln(\tau_1), \ln(\tau_1), \Omega_1, \Omega_2 \rangle$ instead of cell properties directly.

4.4.5 Flow Cytometry

Algae culture samples were prepared for the flow cytometer by adding 930 µL of sodium phosphate buffer (40 mM, pH = 5.16), 50 µL CountBrite Absolute counting bead solution (Life Technologies), and 1 µL BODIPY 505/515 stock solution (1 mg/mL in HPLC-grade dimethyl sulfoxide, stored frozen in 10 µL aliquots) to 10 µL of algae culture. Unstained controls were also included, and stained samples were covered in foil to minimize photo bleaching.

We used the LSR II flow cytometer, (BD Biosciences, San Jose, CA), with 488 nm and 405 nm lasers to measure forward scatter area, side scatter, FITC (530±15 nm excited at 488 nm), PE-Cy7 (780±30 nm excited at 488 nm), PerCP-Cy5.5 (695±20 nm excited at 488 nm), Pacific Blue (450±25 nm excited at 355 nm), and AmCyan (525±25 nm excited at 355 nm). This followed the protocol used by (45), originally developed by (46).

Following the same protocols, we gated the event samples to separate counting beads, cell, and debris events. Cell events were gated using the chlorophyll channels as shown in Figure 4.2. Counting beads were gated as events with Pacific-Blue > 2000 FIU and AmCyan-A > 200 FIU. See Table 4.1 for gating cytograms and voltages.

4.5 Results and Discussion

We grew *Chlamydomonas reinhardtii* cultures, and induced lipid accumulation by resuspending cells in nitrogen-limited TAP media (along with controls suspended in nitrogen-replete media). We measured lipid accumulation by conducting flow cytometry to measure BODIPY fluorescence (46). This protocol has been used to measure lipid accumulation in nitrogen-starved (and replete) *Chlorella vulgaris*, and

Bono et al. found BODIPY fluorescence correlated strongly with quantitative TAG measurements from ^{1}H-NMR spectroscopy (45). Gating cells based on chlorophyll intensity allows the inclusion of a bead standard to verify instrument consistency (Figure 4.2, Table 4.1). Our nitrogen-starved *Chlamydomonas reinhardtii* exhibit increases in fluorescence intensity as compared to their nitrogen-replete counterparts (Figure 4.3), indicating successful induction of lipid accumulation.

On *Chlamydomonas reinhardtii* cells from the same cultures on the same days, we measured single-cell ROT spectra in both nitrogen-starved and nitrogen-replete conditions by applying rotating electric fields at frequencies ranging 1.125 kHz-62.5 MHz and capturing cell rotation from video microscopy. We estimated ROT rates via automated image processing. These rotation rates are proportional to the imaginary part of the complex Clausius-Mossotti factor, which depends on the applied field frequency as well as the algae cell radius, conductivity and permittivity of the cytoplasm, conductivity and permittivity of the suspending medium, and specific capacitance and conductance of the shell (lumped effect of cell wall and cell membrane). By using least trimmed squares, we robustly fit curves parameterized by two peak locations (of $1/\tau_1$ and $1/\tau_2$) and two peak amplitudes ($\Omega_1/2$ and $\Omega_2/2$). (See Figures 4.4 and 4.5.)

Ratios of the fit parameters (τ_1, τ_2, Ω_1, and Ω_2) appear in approximate equations of cell properties, and are either proportional to cytoplasmic conductivity (σ_c) (eqns. (16)(17)) or a sum involving cytoplasmic permittivity ($\varepsilon_c + 2\varepsilon_m$) (eqns. (18)(19)). All four relaxation ratios significantly negatively correlated ($p<0.05$, F-test) with BOIDPY fluorescence (Figure 4.6). Parameter values are reported in Table 4.2. The latter two

correlations indicate that cytoplasmic permittivity may not be constant through lipid accumulation. Thus, the approximate cytoplasmic conductivity of nitrogen-starved *Chlamydomonas reinhardtii* from measured upper crossover frequency (26,27) via the equation of (32) may not be valid with an assumption of constant cytoplasmic permittivity.

The measured relaxation ratios also depend on trans-shell specific conductance and shell specific capacitance, which may also change as a consequence of nitrogen starvation. The lower crossover frequency (which occurs approximately at $1/\tau_1$) of *Chlorella* has been shown to change with increasing lipid content (29,30). Also, the estimated cell membrane and cell wall properties of *Chlorella protothecoides* changed with the lipid accumulation prompted by a change from autotrophic to heterotrophic growth condition (37). Finally, DC-iDEP trapping revealed that dead *Selenastrum capricornutum* had an increased transmembrane conductance associated with a compromised membrane (47). It is possible that shell capacitance and conductance both change with nitrogen starvation in a way that causes the ROT parameter ratios to correlate with lipid accumulation.

In order to fully resolve the change in dielectric properties of our *Chlamydomonas reinhardtii*, we must make an additional measurement beyond a ROT spectrum. A combined single-cell DEP crossover frequency and ROT spectrum measurement as proposed by (39) would resolve this unknown. Additional future work could tie single-cell electrical properties to single-cell fluorescence measurements by taking a BODIPY fluorescence image prior to collecting ROT spectra, and using image processing to extract single-cell BODIPY fluorescence. Thus, heterogeneity of

electrical properties within algae populations may be tied to heterogeneity of accumulated lipids.

Although we took care to terminate the cable-device connections with 50-Ohm resistors, parasitic impedance could have had an effect on the applied field at high frequencies. A lower-than-modeled field strength at high frequencies could have diminished the apparent amplitude of Ω_2 and slightly shifted the apparent location, $1/\tau_2$. To understand the scale of this potential problem, we measured the forward reflection coefficient of a single port in our device (with its adjacent 50-Ohm resistor), and we found the magnitude of the reflection coefficient, |Γ|, to be 0.08 at 65 MHz, compared to <10^-6 at 100 kHz, indicating the presence of high-frequency attenuation. Furthermore, the maximum applied field frequency, 62.5 MHz, is approaching the bandwidth of the function generator, 125 MHz, which would have a similar effect on the apparent Ω_2 and τ_2.

4.6 Conclusion

Electrokinetics show promise for measurement of algae based of their changes in lipid content. Previously, DEP measurements required an assumption of constant cytoplasmic permittivity in order to estimate the cytoplasmic conductivity. Here, electrorotation spectra provided single cell measurements of the effect of lipid accumulation on *Chlamydomonas reinhardtii*. Ratios of these ROT spectrum parameters are proportional to either cytoplasmic conductivity or a sum involving cytoplasmic permittivity. All four parameter ratios significantly ($p<0.05$, F-test) negatively correlated with lipid content (measured by median BODIPY fluorescence

of cells from the same condition), indicating that both cytoplasmic conductivity and permittivity may decrease through lipid accumulation.

4.7 Acknowledgements

The authors would like to thank Dr. Michael Bono, Jr. for his detailed protocols on algae culture, BODIPY staining, and flow cytometry. The authors would also like to thank Dr. Beth Ahner and Dr. Lubna Richter for consultations on algae culture and strain selection. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. (DGE-1144153) (TBL). The authors would like to thank Cornell University's Engineering Learning Initiatives (WWS) for financial support. This work was performed in part at the Cornell NanoScale Facility, a member of the National Nanotechnology Coordinated Infrastructure (NNCI), which is supported by the National Science Foundation (Grant ECCS-1542081). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

4.8 Figures, Tables, and their Legends

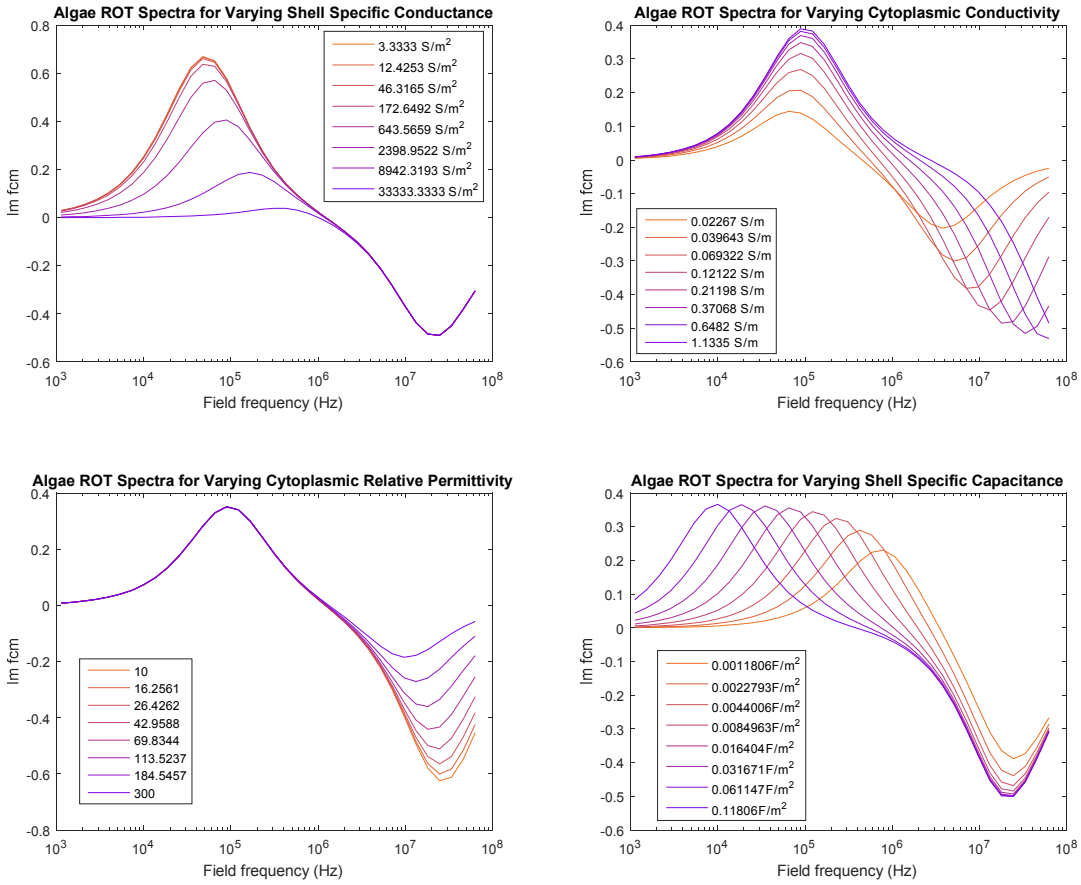

Figure 4.1 – Simulated spectra depicting expected ROT response (imaginary part of the Clausius–Mossotti factor vs. applied field frequency) for a two shell model varying cellular parameters of shell specific conductance, specific capacitance, cytoplasmic conductivity, and cytoplasmic permittivity. Parameters ranges were selected on a log scale around the expected values for nitrogen-replete *Chlamydamonas reinhardtii*.

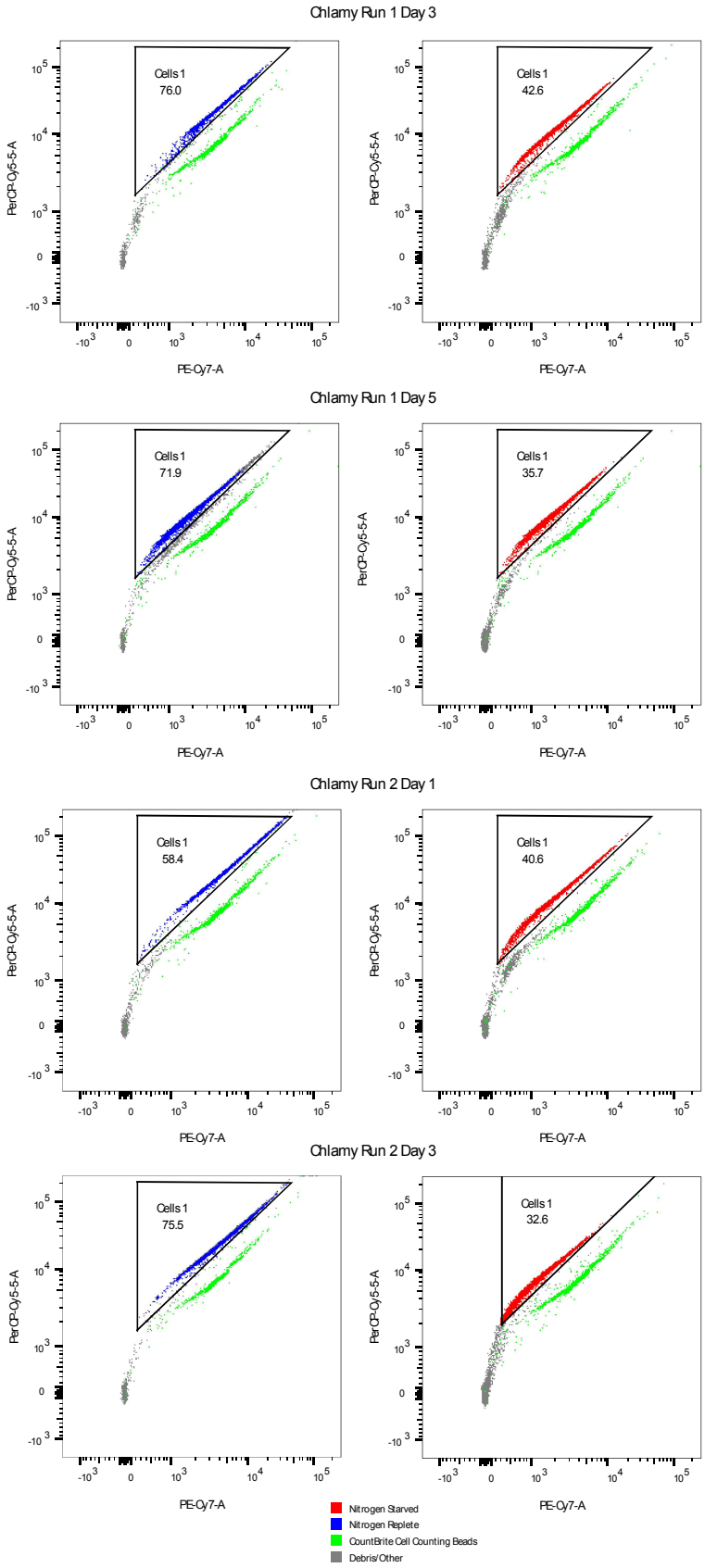
Chlamy Run 1 Day 3
Cells 1
76.0
Cells 1
42.6
Chlamy Run 1 Day 5
Cells 1
71.9
Cells 1
35.7
Chlamy Run 2 Day 1
Cells 1
58.4
Cells 1
40.6
Chlamy Run 2 Day 3
Cells 1
75.5
Cells 1
32.6
PerCP-Cy5-5-A
PE-Cy7-A
Nitrogen Starved
Nitrogen Replete
CountBrite Cell Counting Beads
Debris/Other

Figure 4.2 – Algae flow cytometry events were gated based on chlorophyll auto fluorescence, as described in the Materials and Methods section. Nitrogen-replete cells are shown gated in blue, nitrogen-starved cells are shown gated in red, counting beads are shown to be a light green, and the rest (likely cell debris)—are shown in gray. The numbers inside of the gates are the percentage gated inside. Nitrogen-starved cells generally had a lower cell density, which resulted in overall smaller percentages relative to debris and counting beads.

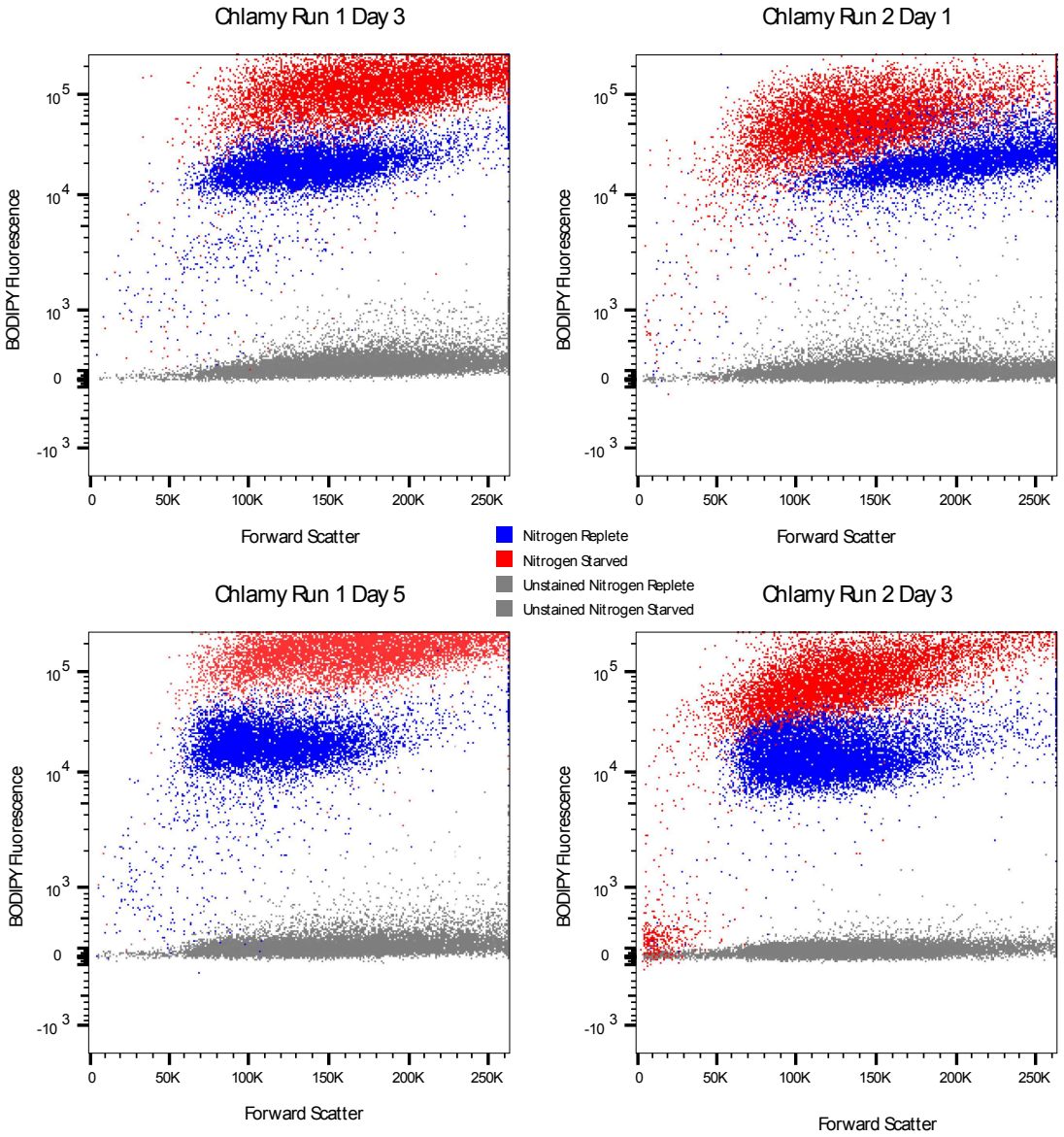

Figure 4.3 – Scatterplots of BODIPY fluorescence, a stain for lipids in algae, of nitrogen-replete (blue) and nitrogen-starved cells (red) plotted against forward scatter for after different runs and nitrogen starvation times. For reference, grey unstained cells also appear on the cytograms. Upward shifts in fluorescence appeared from Day 3 to Day 5 for our first run, and then from Day 1 to Day 3 for our second run, consistent with our expectations for increased lipids with increased time of nitrogen starvation.

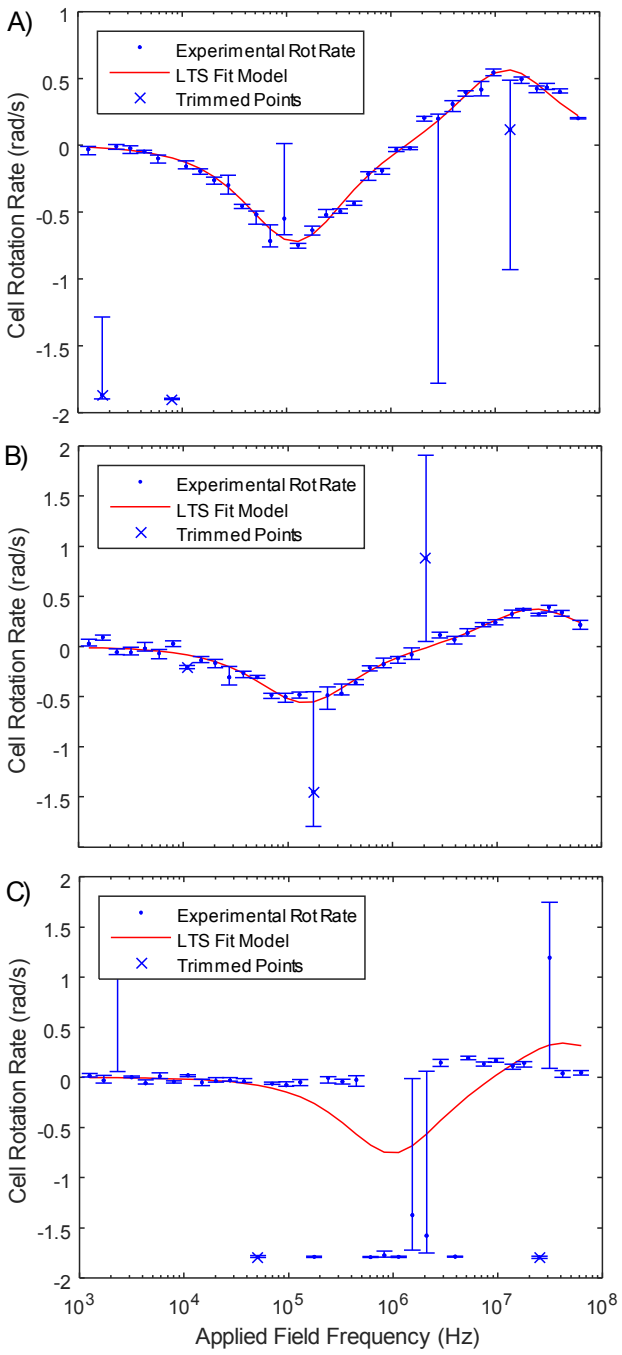

Figure 4.4 – Experimental electrorotation spectra of *Chlamydomonas reinhardtii* with least trimmed squares fits. Bars represent 95% confidence intervals of the median ROT rate in the window over which the frequency was applied. A) is a nitrogen-starved cell, B) is a nitrogen-replete cell. The level of noise and outliers of A) and B) are typical for most (~95%) measured ROT spectra. C) is a cell in which the image processing and curve fit experienced a spectacular failure, and the number of outliers exceeded the number of trimmed points. (These events are rare, below 5% of cells, and these data have a negligible effect on the later correlations due to the use of robust regression.)

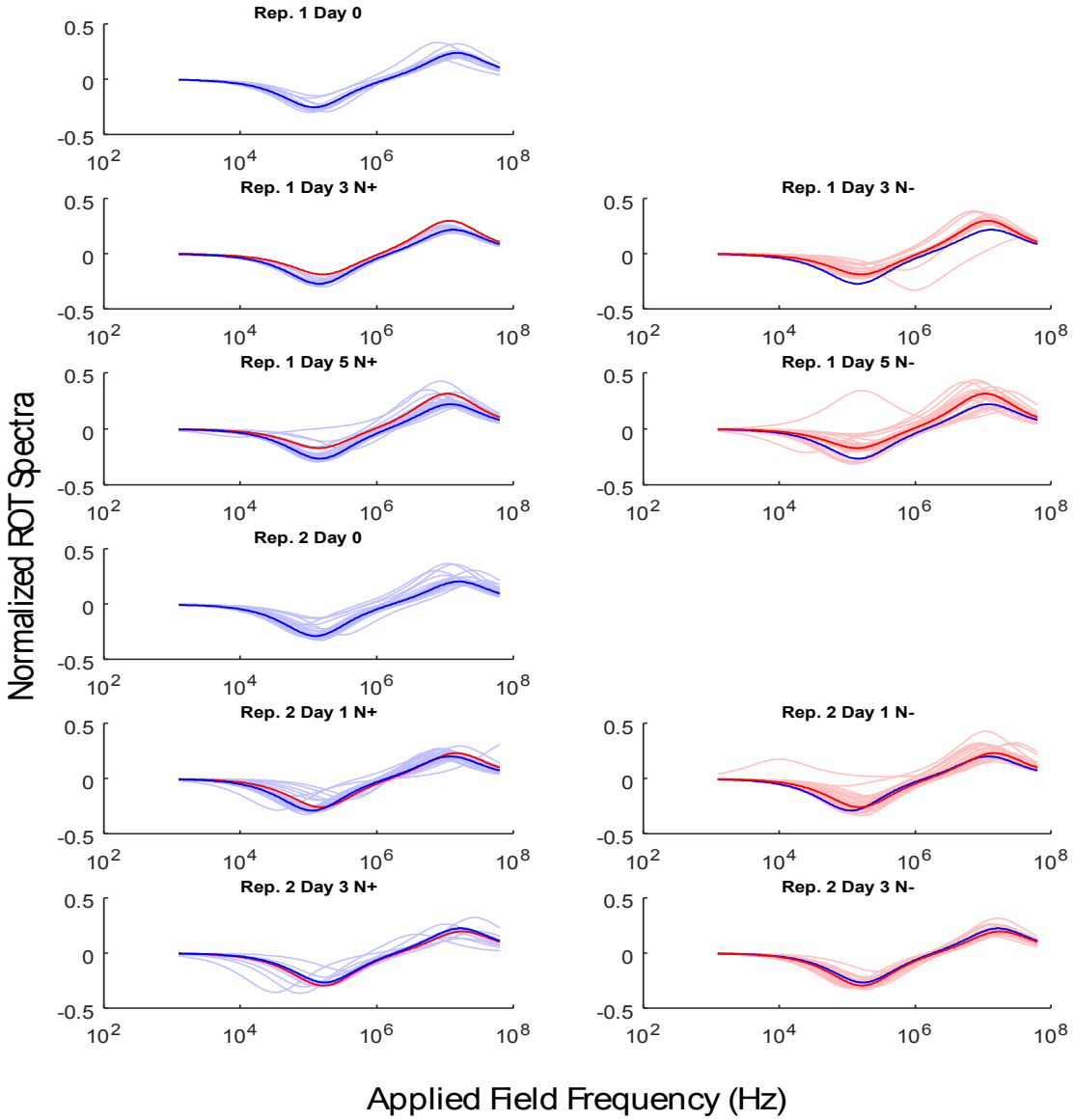

Figure 4.5 – Normalized fit electrorotation spectra of *Chlamydomonas reinhardtii* for each replicate/day and growth media condition: nitrogen-replete (left, blue) or nitrogen-starved (right, red). Spectra were fit with for two peak locations ($1/\tau_1$, $1/\tau_2$) and two peak amplitudes ($\Omega_1/2, \Omega_2/2$) as predicted by a single shell model. Rotation rates (y-axis) were normalized by $|\Omega_1| + |\Omega_2|$. Faint lines represent individual cell spectra. Bold lines indicate a representative curve generated from the median of fit parameters for that day and condition. Representative curves from the opposite condition are included for reference. Note that the amplitude of the high frequency peak of the nitrogen-starved algae is typically larger than that of their nitrogen-replete counterparts.

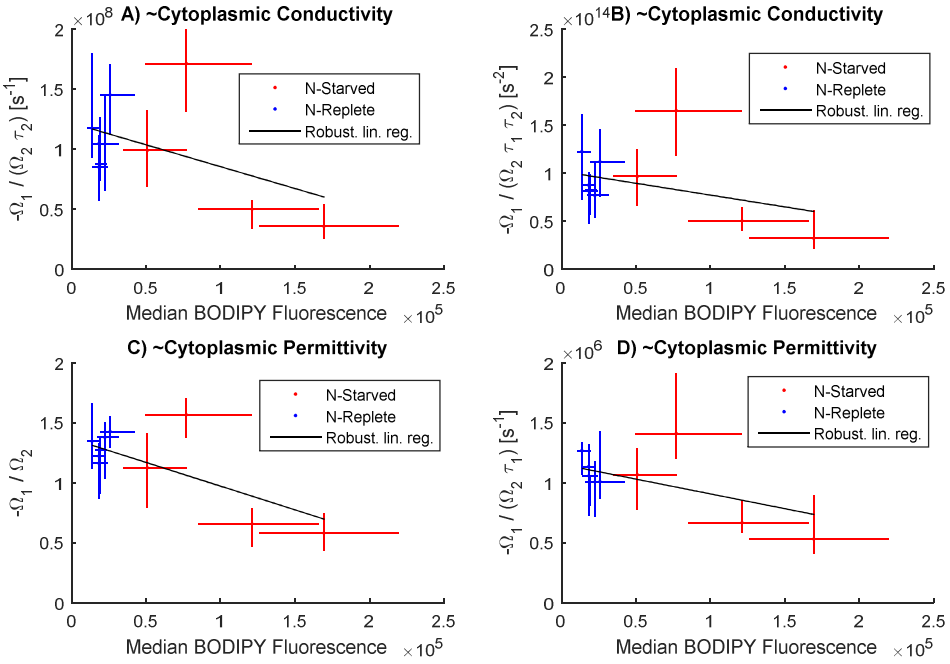

Figure 4.6 – Electrorotation spectrum parameter ratios versus median BODIPY fluorescence from corresponding *Chlamydomonas reinhardtii* cultures as measured by flow cytometry. Bars represent 25-75 percentiles. Individual cell electrical properties versus the sample corresponding median BODIPY fluorescence was fit via robust linear regression. Equations for fits are A) y=-3.64e+02x+1.22e+08, B) y=-2.46e+08x+1.02e+14, C) y=-3.95e-06x+1.37e+00, D) y=-2.46e+00x+1.15e+06. For regression lines, y is in units of: s^{-1} (A), s^{-2} (B), dimensionless (C) and s^{-1} (D) and x must be in fluorescence intensity units as put out by the LSRII flow cytometer. All models are significant at $p<0.05$ (F-test against null of y=constant).

Table 4.1 – Flow cytometry channels, filters, voltages, and use in the flow cytometry experiments to determine lipid content of Chlamydomonas reinhardtii.

Detector	Excitation (nm)	Emission (nm)	Voltage (V)	Use in Study
Forward Scatter (FSC)	-	-	451	Size measurement, preliminary gating
Side Scatter (SSC)	-	-	286	Size measurement, preliminary gating
FITC	488	530 ± 15	302	BODIPY Fluorescence
PE-Cy7	488	780 ± 30	302	Infrared, cell gating x-axis
PerCP-C5.5	488	695 ± 20	302	Far red, cell gating, y-axis
Pacific Blue	405	450 ± 25	302	Counting bead gating
AmCyan	405	525 ± 25	302	Counting bead gating

Table 4.2 – Medians, (25-75 percentiles), and {95% bootstrapped confidence intervals on median} of electrorotation spectrum parameter ratios (equations (16)(17)(18)(19)) and flow-cytometry-measured BODIPY fluorescence of *Chlamydomonas reinhardtii* algae cells. Cell radii were measured manually from cell images.

TAP N+/-	Rep.	Days Since Resus.	N_{ROT}	$-\frac{\Omega_1}{\Omega_2\,\tau_2}$ [ms^{-1}]	$-\frac{\Omega_1}{\tau_2\,\tau_1\,\Omega_2}$ [ms^{-2}]	$-\frac{\Omega_1}{\Omega_2}$ []	$-\frac{\Omega_1}{\tau_1\Omega_2}$ [ms^{-1}]	Radius (μm)	Algae BODIPY Fluorescence	Bead Standard Fluorescence
+	1	0	13	87 (74 - 126) { 76 - 124}	82 (57 - 98) { 59 - 97}	1.27 (0.91 - 1.38) {0.91 - 1.37}	1.00 (0.81 - 1.05) {0.86 - 1.05}	5.1 (4.8 - 5.4) {4.8 - 5.4}	18941 (15893- 23041)	63510 (61680-67010)
+	1	3	15	105 (66 - 111) { 62 - 110}	88 (68 - 99) { 68 - 97}	1.22 (1.05 - 1.29) {1.05 - 1.27}	1.13 (0.96 - 1.21) {0.95 - 1.21}	4.6 (4.3 - 5.7) {4.2 - 5.6}	18243 (14891- 22662)	65078 (62741-69170)
+	1	5	27	85 (57 - 104) { 60 - 101}	82 (47 - 101) { 56 - 97}	1.16 (0.87 - 1.33) {0.94 - 1.27}	1.06 (0.72 - 1.31) {0.83 - 1.18}	4.5 (4.0 - 5.6) {4.1 - 5.5}	18286 (14250- 23819)	64289 (61831-68498)
+	2	0	34	145 (112 - 170) {117 - 160}	112 (75 - 144) { 83 - 130}	1.43 (1.29 - 1.55) {1.34 - 1.50}	1.01 (0.86 - 1.42) {0.92 - 1.30}	5.7 (5.0 - 7.1) {5.2 - 6.8}	25824 (18941- 41833)	66684 (64133-71926)
+	2	1	30	104 (65 - 144) { 71 - 133}	77 (54 - 109) { 55 - 99}	1.38 (1.03 - 1.50) {1.14 - 1.46}	1.01 (0.72 - 1.17) {0.78 - 1.09}	6.4 (5.6 - 7.4) {5.8 - 7.3}	22342 (17125- 31263)	66522 (63821-72809)
+	2	3	11	117 (93 - 179) { 88 - 184}	123 (72 - 161) { 63 - 164}	1.34 (1.12 - 1.66) {1.10 - 1.69}	1.27 (1.07 - 1.33) {1.02 - 1.35}	6.0 (5.6 - 6.2) {5.6 - 6.2}	13638 (10483- 18941)	61082 (58606-64761)
-	1	3	20	50 (34 - 57) { 38 - 56}	51 (40 - 63) { 41 - 60}	0.66 (0.47 - 0.78) {0.52 - 0.77}	0.67 (0.58 - 0.84) {0.59 - 0.80}	5.4 (4.6 - 5.7) {4.8 - 5.7}	121220 (126393- 165699)	64761 (62436-69170)
-	1	5	29	36 (25 - 54) { 26 - 43}	32 (21 - 60) { 29 - 50}	0.58 (0.43 - 0.74) {0.47 - 0.69}	0.53 (0.40 - 0.89) {0.42 - 0.71}	4.9 (4.6 - 5.2) {4.7 - 5.1}	169416 (169416 - 219501)	64289 (61831-68498)
-	2	1	40	99 (69 - 132) { 91 - 116}	97 (66 - 124) { 77 - 113}	1.12 (0.79 - 1.40) {0.98 - 1.20}	1.07 (0.78 - 1.28) {0.87 - 1.22}	4.7 (3.9 - 5.4) {4.1 - 5.1}	50900 (34663- 77019)	67999 (65237-73884)
-	2	3	36	171 (131 - 199) {149 - 190}	165 (118 - 208) {141 - 186}	1.57 (1.37 - 1.70) {1.49 - 1.63}	1.41 (1.20 - 1.91) {1.29 - 1.63}	4.4 (3.7 - 5.0) {4.0 - 4.7}	77019 (49680- 120922)	63665 (61231-68331)

4.9 References

1. Mata TM, Martins A a., Caetano NS. Microalgae for biodiesel production and other applications: A review. Renew. Sustain. Energy Rev. 2010;14:217–232. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1364032109001646.

2. Chisti Y. Biodiesel from microalgae beats bioethanol. Trends Biotechnol. 2008;26:126–131. Available at: http://dx.doi.org/10.1016/j.biotechadv.2007.02.001.

3. Fukuda H, Kondo A, Noda H. Biodiesel fuel production by transesterification of oils. J. Biosci. Bioeng. 2001;92:405–416.

4. Johnson MB, Wen Z. Production of biodiesel fuel from the microalga schizochytrium limacinum by direct transesterification of algal biomass. Energy and Fuels 2009;23:5179–5183.

5. Huang G, Chen F, Wei D, Zhang X, Chen G. Biodiesel production by microalgal biotechnology. Appl. Energy 2010;87:38–46. Available at: http://dx.doi.org/10.1016/j.apenergy.2009.06.016.

6. Pruvost J, Van Vooren G, Cogne G, Legrand J. Investigation of biomass and lipids production with Neochloris oleoabundans in photobioreactor. Bioresour. Technol. 2009;100:5988–5995. Available at: http://dx.doi.org/10.1016/j.biortech.2009.06.004.

7. Verthein R, Macedo T De, Alegre RM. INFLUÊNCIA DO TEOR DE NITROGÊNIO NO CULTIVO DE Spirulina maxima 3 – RESULTADOS E DISCUSSÃO 2 – MATERIAL E MÉTODOS. Rev. Lit. Arts Am. 2001;21:183–186.

8. Rodolfi L, Zittelli GC, Bassi N, Padovani G, Biondi N, Bonini G, Tredici MR. Microalgae for oil: Strain selection, induction of lipid synthesis and outdoor mass cultivation in a low-cost photobioreactor. Biotechnol. Bioeng. 2009;102:100–112.

9. Stephenson AL, Kazamia E, Dennis JS, Howe CJ, Scott S a., Smith AG. Life-cycle assessment of potential algal biodiesel production in the united kingdom: A comparison of raceways and air-lift tubular bioreactors. Energy and Fuels 2010;24:4062–4077.

10. Hu Q, Sommerfeld M, Jarvis E, Ghirardi M, Posewitz M, Seibert M, Darzins A. Microalgal triacylglycerols as feedstocks for biofuel production: Perspectives and advances. Plant J. 2008;54:621–639.

11. Hoffmann M, Marxen K, Schulz R, Vanselow KH. TFA and EPA productivities of Nannochloropsis salina influenced by temperature and nitrate stimuli in turbidostatic controlled experiments. Mar. Drugs 2010;8:2526–2545.

12. Gao C, Xiong W, Zhang Y, Yuan W, Wu Q. Rapid quantitation of lipid in microalgae by time-domain nuclear magnetic resonance. J. Microbiol. Methods 2008;75:437–440. Available at: http://dx.doi.org/10.1016/j.mimet.2008.07.019.

13. Gardner WS, Frez W a., Cichocki E a., Parrish CC. Micromethod for lipids in aquatic invertebrates. Limnol. Oceanogr. 1985;30:1099–1105.

14. Lee JY, Yoo C, Jun SY, Ahn CY, Oh HM. Comparison of several methods for effective lipid extraction from microalgae. Bioresour. Technol. 2010;101:S75–S77. Available at: http://dx.doi.org/10.1016/j.biortech.2009.03.058.

15. Jones J, Manning S, Montoya M, Keller K, Poenie M. Extraction of algal lipids and their analysis by HPLC and mass spectrometry. JAOCS, J. Am. Oil Chem. Soc. 2012;89:1371–1381.

16. Archanaa S, Moise S, Suraishkumar GK. Chlorophyll interference in microalgal lipid quantification through the Bligh and Dyer method. Biomass and Bioenergy 2012;46:805–808.

17. Chen W, Zhang C, Song L, Sommerfeld M, Hu Q. A high throughput Nile red method for quantitative measurement of neutral lipids in microalgae. J. Microbiol. Methods 2009;77:41–47. Available at: http://dx.doi.org/10.1016/j.mimet.2009.01.001.

18. Govender T, Ramanna L, Rawat I, Bux F. BODIPY staining, an alternative to the Nile Red fluorescence method for the evaluation of intracellular lipids in microalgae. Bioresour. Technol. 2012;114:507–511. Available at: http://dx.doi.org/10.1016/j.biortech.2012.03.024.

19. Davey PT, Hiscox WC, Lucker BF, O'Fallon J V., Chen S, Helms GL. Rapid triacylglyceride detection and quantification in live micro-algal cultures via liquid state 1H NMR. Algal Res. 2012;1:166–175. Available at: http://dx.doi.org/10.1016/j.algal.2012.07.003.

20. Bono MS. DIELECTRIC MEASUREMENT OF ALGAL LIPID CONTENT FOR BIODIESEL PRODUCTION. 2015.

21. Kirby BJ. Micro- and nanoscale fluid mechanics : Transport in Microfluidic Devices. 1st ed. Cambridge: Cambridge University Press; 2010. Available at: http://www.kirbyresearch.com/index.cfm/wrap/textbook/microfluidicsnanofluidics.html.

22. Jones TB. Electromechanics of Particles. Cambridge: Cambridge University Press; 1995. Available at: http://dx.doi.org/10.1017/CBO9780511574498. Accessed February 18, 2014.

23. Morgan H, Green NG. AC Electrokinetics: Colloids and Nanoparticles. Hertfordshire, England: Research Studies Press Ltd.; 2003.

24. Foster KR, Sauer F a., Schwan HP. Electrorotation and levitation of cells and colloidal particles. Biophys. J. 1992;63:180–90. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0006349592815886. Accessed January 6, 2014.

25. Pethig R. Dielectrophoresis: Status of the theory, technology, and applications. Biomicrofluidics 2010;4.

26. Hadady H, Wong JJ, Hiibel SR, Redelman D, Geiger EJ. High frequency dielectrophoretic response of microalgae over time. Electrophoresis 2014:1–18. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25229637.

27. Michael K a., Hiibel SR, Geiger EJ. Dependence of the dielectrophoretic upper crossover frequency on the lipid content of microalgal cells. Algal Res. 2014;6:17–21. Available at: http://dx.doi.org/10.1016/j.algal.2014.08.004.

28. Bono MS, Ahner B a, Kirby BJ. Detection of algal lipid accumulation due to nitrogen limitation via dielectric spectroscopy of Chlamydomonas reinhardtii suspensions in a coaxial transmission line sample cell. Bioresour. Technol. 2013;143:623–31. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23845710. Accessed January 21, 2014.

29. Deng YL, Kuo MY, Juang YJ. Development of flow through dielectrophoresis microfluidic chips for biofuel production: Sorting and detection of microalgae with different lipid contents. Biomicrofluidics 2014;8. Available at: http://dx.doi.org/10.1063/1.4903942.

30. Deng YL, Chang JS, Juang YJ. Separation of microalgae with different lipid contents by dielectrophoresis. Bioresour. Technol. 2013;135:137–141. Available at: http://dx.doi.org/10.1016/j.biortech.2012.11.046.

31. Schor AR, Buie CR. Non-Invasive Sorting of Lipid Producing Microalgae With Dielectrophoresis Using Microelectrodes. In: Vol. 9 Micro- Nano-Systems Eng. Packag. Parts A B. ASME; 2012. p 701. Available at: http://proceedings.asmedigitalcollection.asme.org/proceeding.aspx?doi=10.1115/IMECE2012-88317.

32. Broche LM, Labeed FH, Hughes MP. Extraction of dielectric properties of multiple populations from dielectrophoretic collection spectrum data. Phys. Med. Biol. 2005;50:2267–74. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-20044380970&partnerID=tZOtx3y1.

33. Hawkins BG, Huang C, Arasanipalai S, Kirby BJ. Automated Dielectrophoretic Characterization of Mycobacterium smegmatis. Anal. Chem. 2011;83:3507–3515. Available at: http://pubs.acs.org/doi/abs/10.1021/ac2002017.

34. Huang Y, Pethig R. Electrode design for negative dielectrophoresis. Meas. Sci. Technol 1991;1142:2–7.

35. Rohani A, Varhue W, Su Y, Swami NS. Electrical tweezer for highly parallelized electrorotation measurements over a wide frequency bandwidth. Electrophoresis 2014;35:1795–802. Available at: http://doi.wiley.com/10.1002/elps.201400021\nhttp://www.ncbi.nlm.nih.gov/pubmed/24668830.

36. Muller T, Schnelle T, Fuhr G. Dielectric single cell spectra in snow algae. Polar Biol 1998;20:303–310.

37. Wu Y, Huang C, Wang L, Miao X, Xing W, Cheng J. Electrokinetic system to determine differences of electrorotation and traveling-wave electrophoresis between autotrophic and heterotrophic algal cells. Colloids Surfaces A Physicochem. Eng. Asp. 2005;262:57–64.

38. Kirby BJ. Micro- and Nanoscale Fluid Mechanics. New York: Cambridge University Press; 2013.

39. Chan KL, Gascoyne PR. C, Becker FF, Pethig R. Electrorotation of liposomes: verification of dielectric multi-shell model for cells. Biochim. Biophys. Acta - Lipids Lipid Metab. 1997;1349:182–196. Available at: http://doi.wiley.com/10.1002/rcm.1690.

40. Deng X, Fei X, Li Y. The effects of nutritional restriction on neutral lipid accumulation in Chlamydomonas and Chlorella. African J. Microbiol. Res. 2011;5:260–270. Available at: http://www.academicjournals.org/ajmr.

41. Gorman DS, Levine RP. Cytochrome f and plastocyanin: their sequence in the photosynthetic electron transport chain of Chlamydomonas reinhardi. Proc. Natl. Acad. Sci. U. S. A. 1965;54:1665–1669.

42. Malmberg CG, Maryott a. a. Dielectric constants of aqueous solutions of dextrose and sucrose. J. Res. Natl. Bur. Stand. (1934). 1950;45:299.

43. Witman GB, Carlson K, Berliner J, Rosenbaum JL. CHLAMYDOMONAS FLAGELLA: I. Isolation and Electrophoretic Analysis of Microtubules, Matrix, Membranes, and Mastigonemes. J. Cell Biol. 1972;54:507–539. Available at: http://www.jcb.org/cgi/doi/10.1083/jcb.54.3.507.

44. Lannin TB, Su W-W, Gruber C, Cardle I, Huang C, Thege FI, Kirby BJ. Automated electrorotation shows the electrokinetic separation of pancreatic cancer cells is robust to acquired chemotherapy resistance, serum starvation, and epithelial-to-mesenchymal transition. Biomicrofluidics (Submitted) 2016.

45. Bono MS, Garcia RD, Sri-Jayantha D V., Ahner B a., Kirby BJ. Measurement of lipid accumulation in Chlorella vulgaris via flow cytometry and liquid-state 1H NMR spectroscopy for development of an NMR-traceable flow cytometry protocol. PLoS One 2015;10:1–18.

46. Cirulis JT, Strasser BC, Scott J a., Ross GM. Optimization of staining conditions for microalgae with three lipophilic dyes to reduce precipitation and fluorescence variability. Cytom. Part A 2012;81 A:618–626.

47. Gallo-Villanueva RC, Jesús-Pérez NM, Martínez-López JI, Pacheco A, Lapizco-Encinas BH. Assessment of microalgae viability employing insulator-based dielectrophoresis. Microfluid. Nanofluidics 2011;10:1305–1315. Available at: http://link.springer.com/10.1007/s10404-010-0764-3.

CHAPTER 5

Conclusions

5.1 Conclusions

In this dissertation, I have developed two tools (a set of machine learning algorithms for cell classifications and an electrorotation apparatus with accompanying image processing software for automated analysis) for three characterizations of cells (rapid location and identification of captured cancer cells, electrical property measurement of pancreatic cancer cells, and electrical property measurement of algae cells). In Chapter 2, I showed that machine learning algorithms can classify image patches of fluorescently-stained cells from cancer patient blood, even in a low signal-to-noise regime where disagreements exist among the manual training experts. These tools helped the progress of a phase II clinical trial (TAXYNERGY) (1), helped verify a probabilistic capture model for cells in the GEDI device(2), helped verify the existence of early-disseminating pancreatic cancer cells in high-risk populations (3), and helped profile captured patient pancreatic circulating cells (4).

I see two future directions for this work. First, new image-specific machine learning algorithms have emerged, (e.g. convolutional neural networks), which could lead to better classification performance than the four general-purpose machine learning methods I employed. It would be worthwhile to test their performance. Second, the part of my tool that extracted features of the cells from the image patches (e.g. area of nucleus, brightness of one stain, correlation between two stains) has value beyond preparing data for machine learning. Adding a user interface to this tool in

which the measured cell attributes could be gated and manipulated like in flow cytograms could aid the rapid analysis of GEDI images.

In Chapter 3, I developed automated electrorotation to measure dielectric properties of pancreatic cancer cells under three perturbations: acquired resistance to gemcitabine, serum starvation, and epithelial-to-mesenchymal transition. Regardless of treatment, the crossover frequency of all cancer cells remained significantly higher than that of white blood cells isolated from a healthy donor.

Interesting potential future work could involve measuring ROT spectra of different pancreatic cancer cell lines, particularly different gemcitabine-resistant sub clones, to see if a low crossover frequency is retained through more cellular transformations. It would also be valuable to verify the computed crossover frequencies against DEP trapping experiments to confirm the model. Finally, because single ROT spectra are measured, it would be interesting to stain the cells for surface markers of interest, capture a fluorescence image, and tie specific markers to dielectric characteristics of the cells.

In Chapter 4, I employed my automated electrorotation system to study changes to the (approximate) cytoplasmic conductivity and permittivity of nitrogen-starved Chlamydomonas reinhardtii. With an assumption of either constant shell capacitance or conductance, these properties had a significant negative correlation with median BODIPY fluorescence as measured by flow cytometry, as would be predicted by high-conductivity, high-permittivity salt water being displaced by low-conductivity, low-permittivity lipids.

Future work could involve ROT measurements to inform the operation of DEP devices that separate cells based on lipid content. Such devices could be used to select for mutant strains of high-lipid-yield algae. Additional future work could involve simultaneous BODIPY fluorescence and ROT experiments, tying single cell electrical properties with single cell fluorescence measurements.

Automated image processing has been central to all of my thesis work. It enables quantitative extraction of data that otherwise would have only been quantitative. It can increase data processing speed by orders of magnitude, and make it perfectly repeatable. It also enables a wider range of experiments to be conducted (and even imagined) with a smaller prerequisite of tools, because we can expect that meaningful, quantitative results can be extracted with a camera and a computer.

5.2 References

1. Tagawa ST, Galletti G, Antonarakis ES, Tasaki S, Gjyrezi A, Worroll D, Portella L, Kirby BJ, Stewart J, Zaher A, Saad F, Vanhuyse M, Suri S, Lannin TB, Gruber C, Pratt E, North S, Eisenberger M a., Nanus DM, Giannakakou P. 2563 Screening and baseline analysis of circulating tumor cell (CTC) counts and androgen receptor (AR) localization with clinical characteristics of men with metastatic castration-resistant prostate cancer (mCRPC) in TAXYNERGY. Eur. J. Cancer 2015;51:S498–S499. Available at: http://meeting.ascopubs.org/cgi/content/abstract/33/15_suppl/5031\nhttp://linkinghub.elsevier.com/retrieve/pii/S095980491631382X.

2. Smith JP, Lannin TB, Syed YA, Santana SM, Kirby BJ. Parametric control of collision rates and capture rates in geometrically enhanced differential immunocapture (GEDI) microfluidic devices for rare cell capture. Biomed. Microdevices 2013. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24078270. Accessed November 13, 2013.

3. Rhim AD, Thege FI, Santana SM, Lannin TB, Saha TN, Tsai S, Maggs LR, Kochman ML, Ginsberg GG, Lieb JG, Chandrasekhara V, Drebin J a., Ahmad N,

Yang YX, Kirby BJ, Stanger BZ. Detection of circulating pancreas epithelial cells in patients with pancreatic cystic lesions. Gastroenterology 2014;146:647–651. Available at: http://dx.doi.org/10.1053/j.gastro.2013.12.007.

4. Thege FI, Lannin TB, Saha TN, Tsai S, Kochman ML, Hollingsworth M a, Rhim AD, Kirby BJ. Microfluidic immunocapture of circulating pancreatic cells using parallel EpCAM and MUC1 capture: characterization, optimization and downstream analysis. Lab Chip 2014;14:1775–84. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24681997. Accessed October 4, 2014.